A Survey of Social Network Analysis Techniques and their Applications to Socially Aware Networking

著者（英）	Sho TSUGAWA
journal or publication title	IEICE transactions on communications
volume	E102.B
number	1
page range	17-39
year	2019-02
権利	Copyright(C)2019 The Institute of Electronics, Information and Communication Engineers
URL	http://hdl.handle.net/2241/00154602

doi: 10.1587/transcom.2017EBI0003

INVITED SURVEY PAPER

A Survey of Social Network Analysis Techniques and their Applications to Socially Aware Networking

Sho TSUGAWA[†a)], ***Member***

SUMMARY Socially aware networking is an emerging research field that aims to improve the current networking technologies and realize novel network services by applying social network analysis (SNA) techniques. Conducting socially aware networking studies requires knowledge of both SNA and communication networking, but it is not easy for communication networking researchers who are unfamiliar with SNA to obtain comprehensive knowledge of SNA due to its interdisciplinary nature. This paper therefore aims to fill the knowledge gap for networking researchers who are interested in socially aware networking but are not familiar with SNA. This paper surveys three types of important SNA techniques for socially aware networking: identification of influential nodes, link prediction, and community detection. Then, this paper introduces how SNA techniques are used in socially aware networking and discusses research trends in socially aware networking.
key words: *social network analysis, socially aware networking, influence, link prediction, community*

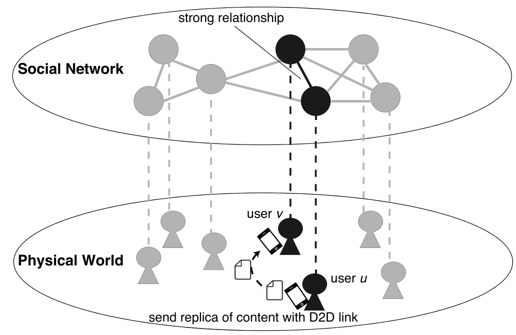


Fig. 1 An example scenario of socially aware networking: Users u and v are close friends and have similar interests. They share a content cache through a D2D link. If user v wants to access content that is already shared by user u, user v can immediately access it.

1. Introduction

Thanks to the proliferation of various network services, fine-grained and large-scale log data of communication among individuals has become available, driving studies on social network analysis (SNA) [1–5]. In SNA, social networks (also called social graphs), in which individuals are represented by nodes and social ties among them are represented by links, are constructed from various data related to human communication [1, 2, 5, 6]. The constructed social networks are then analyzed with the aim of understanding complex social phenomena that involve interactions among a large number of people. Various types of communication data, such as records of mobile phones [7–9], email [9–11], and activities and relationships on social media [12–14], have been used for SNA. These studies reveal universal characteristics of social networks, such as power-law degree distribution [10], small shortest path length [15], high clustering coefficient [15], and community structure [13, 16].

In the literature, many useful techniques for SNA have been proposed, and those techniques have been applied to various domains. For instance, techniques for identifying influential nodes have been used for viral marketing [17], those for predicting future link formation have been used for user recommendation [18], and those for detecting communities in a social network have been used for predicting information diffusion on social media [19].

More recently, SNA techniques have attracted the attention of researchers in the communication networking community, leading to the emergence of a new research field, *socially aware networking* [20]. There is an increasing need to take human factors into account for the design and control of communication networks [4,20]. To fulfill this need, SNA techniques have been applied to problems in communication networking. An illustrative example scenario of socially aware networking is shown in Fig. 1. Suppose that users u and v are close friends who have similar interests, which can be inferred from a social network, and that user u has a specific content item in his/her mobile device. User u sends a cache (or replica) of the content to user v via device-to-device (D2D) wireless link when they meet each other. Then, user v will be able to use the content whenever desired. Recall that users u and v have similar interests. Therefore, it is expected that user v will access the content, which has already been accessed by user u. In this scenario, a socially aware networking scheme infers content access probability based on a social network and allocates content caches using a D2D link, which should improve the quality of service (QoS) of content delivery and also reduce the amount of mobile network traffic. While this is a very simple example of socially aware networking, there have been many proposals for novel communication technologies using SNA techniques [20].

This paper surveys promising SNA techniques that can be used for socially aware networking, as well as recent research trends in socially aware networking. To conduct research on socially aware networking, knowledge of both SNA and communication networking technologies is nec-

Manuscript received September 29, 2017.
Manuscript revised February 9, 2018.
Manuscript publicized February 21, 2018.
[†]The author is with Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba-shi, 305-8573 Japan.
a) E-mail: s-tugawa@cs.tsukuba.ac.jp
DOI: 10.1587/transcom.2017EBI0003

essary. However, since SNA techniques have been mainly developed in the network science and data mining research communities, SNA has only recently received attention from researchers in the communication networking field. Obtaining comprehensive knowledge of SNA techniques is not an easy task for researchers who are new to this field, since research on SNA is interdisciplinary. Important studies are distributed among several fields, including social science, computer science, physics, and applied mathematics. This paper therefore aims to fill the knowledge gap for networking researchers who are interested in socially aware networking but are not familiar with SNA. For readers who already have a strong background in SNA, this paper aims to introduce socially aware networking and promising application domains for SNA.

There are several surveys and textbooks that introduce SNA and socially aware networking. We introduce them as references for readers who aim to gain further understanding of topics not covered by this paper. The history of SNA in social science is introduced in a survey paper by Borgatti et al. [2] and a textbook by Scott [5]. For the theoretical aspects of social networks and complex networks, good textbooks by Newman [21] and Barabási [22] are available. Although socially aware networking is a new research field, an excellent survey paper has been written by Xia et al. [20]. While Xia et al. [20] aimed for a comprehensive survey of socially aware networking, the present paper aims to be a bridge between SNA and socially aware networking research fields.

The remainder of this paper is organized as follows. Section 2 gives an overview of SNA. Promising SNA techniques for identifying influential nodes, link prediction, and community detection are introduced in Sects. 3, 4, and 5, respectively. These sections introduce conventional techniques that can be used for socially aware networking and also cover the state-of-the-art techniques that have potential applications for socially aware networking. Section 6 introduces socially aware networking studies and explains how SNA techniques are used in these studies. In particular, we will introduce studies on socially aware *routing* and socially aware *caching* schemes. Section 7 discusses promising SNA techniques that can be applied to socially aware networking and also discusses open issues. Finally, Sect. 8 concludes this paper.

2. Overview of Social Network Analysis

This section first gives a taxonomy of techniques used in SNA and discusses important techniques in the context of socially aware networking. Then, we introduce several representations of social networks as well as the notations used in this paper.

2.1 Taxonomy of Social Network Analysis Techniques

Although there are various techniques that are studied in the SNA research community, we can broadly classify these techniques into four categories according to their scope: nodes, links, subgraphs, and entire graphs. Table 1 summarizes some representative techniques studied in the literature. In Table 1, related survey papers are also included. Since the components of a social network are nodes and links, inference of node and link characteristics has been actively studied. The first and second categories include techniques for estimating microscopic-level (i.e., at the level of nodes and links) characteristics of networks. Many social networks have interesting subgraph structures, and therefore the detection of such structures of networks is also actively studied. The third category includes techniques related to mesoscopic-level structures of networks. Nodes and links collectively build the complex structure of a social network, and quantifying such complex structures is an important research topic. The fourth category is related to macroscopic-level characteristics of social networks. Note that this taxonomy does not include modeling network generation [23], information diffusion [24], or node mobility [25], even though these are important topics related to social networks. This paper focuses on *analysis*, rather than *modeling*, of social networks.

Among the various SNA techniques, this paper particularly focuses on *identification of influential nodes*, *link prediction*, and *community detection*. These are fundamental research problems in the SNA research field and crucial techniques for socially aware networking [20]. While other SNA techniques are also interesting and might be applied to socially aware networking in future studies, the introduction of those techniques is beyond the scope of this paper. Readers who are interested in other techniques can refer to the references shown in Table 1.

2.2 Social Network Representation

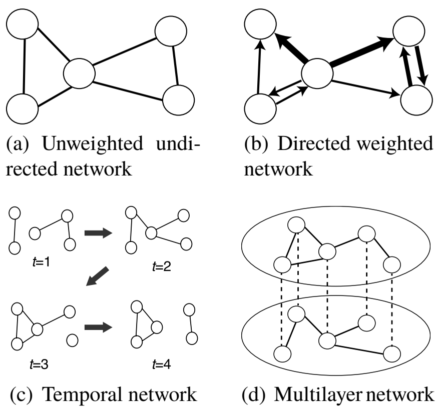
In SNA, individuals and the social relationships between those individuals are represented as a network, in which nodes represent individuals and links represent social ties among them [1,2,5,6]. The representative network is called a social network or a social graph. In this paper, we mainly use the term social network, but we use the terms graph and network interchangeably.

Social networks are constructed from various types of data, with the choice of data depending on the objective of the analysis [5]. Traditional SNA often uses questionnaire data to construct networks of friendships, acquaintance relationships, or trust relationships [1, 5, 37]. Recent SNA studies have constructed social networks from log data of online communications, such as email logs [10,11], records of phone calls [7,8], and activity logs on social media [12,14]. Relational data, such as friendships in social networking services (SNSs) [13] or co-authorship [38] of research papers, are also often used. Log data of offline spatial proximity is also available for SNA [39]. Among various SNA studies, constructing a social network is a common and important step [5].

There are several ways to represent a social network (Fig. 2). Let $G = (V, E)$ be a social network, where V

Table 1 Taxonomy of social network analysis techniques.

Category	Techniques	Related survey papers	Description
Node	**Identification of influential nodes**	[26]	Identifying influential nodes from a network
	Node label estimation	[27]	Determining node label of each node
	Node embedding	[28]	Representing node characteristics as a vector
Link	**Link prediction**	[29, 30]	Predicting future link formation in a network
	Inferring trust	[31]	Inferring trust relationship between two nodes
Subgraph	**Community detection**	[32–34]	Clustering nodes into communities
	Motif detection	[35]	Finding frequently observed subgraph patterns
Entire graph	Quantifying graph characteristics	[36]	Quantifying characteristics of a given network



(a) Unweighted undirected network (b) Directed weighted network (c) Temporal network (d) Multilayer network

Fig. 2 Representations of social networks.

Table 2 Notation used in this paper.

G	a social network
V	set of nodes in graph G
E	set of links in graph G
W	set of link weights in graph G
N	number of nodes in graph G
$\Gamma(v)$	set of nodes adjacent to node v

is a set of nodes representing individuals and E is a set of links representing social ties. Links can be either directed or undirected. Moreover, since social ties have strengths [40], a social network can be represented as a weighted network $G = (V, E, W)$, where W is a set of link weights that characterize the strengths of social ties. The appropriate choice of representation depends on the objective of the SNA and the data used for constructing the social network. When we analyze a friendship network on Facebook, for example, the network can be represented as an unweighted undirected network since friendships have no direction and Facebook friendship data does not offer direct information about tie strengths. When we analyze an email network, in contrast, the network can be represented as a weighted directed network, where link direction represents email direction and link weight represents the frequency of email communication. However, we can intentionally represent the email network as an unweighted graph or an undirected graph by forgetting some information. For example, if we are interested in only strong relationships, each link can represent a particular relationship between a pair of individuals who frequently communicate with each other. If we consider mutual communication, the email network can be an undirected graph, where a link represents the existence of mutual communication. The appropriate social network representation depends on the problem, and so the choice of representation is an important step in SNA.

Other options exist for the representation of social networks: multilayer networks [41, 42] and temporal networks [43, 44] are two notable types. Multilayer networks can represent different types of links, such as email and phone call communications (Fig. 2(d)). Temporal networks can represent the time evolution of networks (Fig. 2(c)). These two types of networks are currently being actively studied and are expected to be powerful tools for representing complex relationships among individuals [41–44]. Although these representations should be useful for various applications, studies on multilayer and temporal networks have not matured. Therefore, in this paper, we mainly introduce studies on single-layer, and non-temporal networks. For more information about temporal networks, please refer to [43–45], and for multilayer networks, please refer to [41, 42].

In what follows, the symbols shown in Table 2 are used in explanations of SNA techniques.

3. Identifying Influential Nodes

One of the most well-studied problems in the SNA research field is the identification of influential nodes in a given social network [17, 26, 46, 47, 47–50]. Influential nodes have also been called central nodes [51], important nodes [52], superspreaders [47], and key players [53]. Although different studies use slightly different definitions [26], in this paper, influential nodes are defined as those that can spread information to many other nodes in a social network. Typical applications of identifying influential nodes are efficient information dissemination, viral marketing, and preventing the spread of rumors and viruses [17, 54].

In socially aware networking schemes, influential node identification is used for identifying nodes that have a higher chance to disseminate information to many other nodes [20]. A more detailed explanation will be given in Sect. 6.

This section first gives a definition of influential node identification problems (Sect. 3.1). Then, several measures and algorithms for the problems will be introduced (Sects. 3.2 and 3.3). We will also briefly introduce other related research topics (Sect. 3.4).

3.1 Problem Formulation

Influential node identification is typically formulated as a problem of estimating the influence of each node. The influence estimation problem is defined as follows.

Problem 1 (Influence Estimation Problem) *Given a social network $G = (V, E, W)$, where V is a set of nodes, E is a set of links, and W is a set of link weights, the aim is to find an influence score $c(v)$ for each node $v \in V$.*

While the influence score $c(v)$ represents the strength of the influence of node v, we are often concerned with relative rankings in terms of influence score $c(v)$, rather than raw score, and we are typically most interested in highly ranked nodes [55, 56].

Influential node identification is also formulated as the problem of detecting a set of influential nodes from a social network. This type of problem is called an influence maximization (IM) problem [17, 57, 58]. This problem was first studied by Domingos and Richardson [57, 58], and Kempe et al. [17] formulated the problem as a combinatorial optimization problem. IM aims to identify a small set of influential nodes ("seed nodes") such that the expected size of the influence cascade triggered by the seed nodes is maximized [17]. Let $G = (V, E, W)$ be a social network, where V is a set of nodes, E is a set of links, and W is a set of link weights representing the probability of influence spread. Let $S \subseteq V$ be a subset of nodes in graph G and $\sigma(S)$ be the expected number of active nodes at the end of the process of an influence cascade model when S is the initial set of active nodes (i.e., the seed nodes). The IM problem is then defined as follows [17].

Problem 2 (Influence Maximization (IM) Problem) *Given a social network G, an integer k, and an influence cascade model, the aim is to find a set of seed nodes S ($S \subseteq V$ with $|S| = k$) such that $\sigma(S)$, which we call the influence spread, is maximized under the given cascade model.*

We introduce measures for the influence estimation problem in Sect. 3.2 and algorithms for the influence maximization problem in Sect. 3.3.

3.2 Measures for Influence Estimation

Various measures for estimating the influence of nodes have been proposed in the literature [26]. This section introduces conventional measures as well as promising, recently proposed ones. Please also refer to the recent survey paper [26] for further information. In what follows, unless explicitly stated otherwise, we assume that a social network G is an unweighted and undirected graph. Most of the measures can be naturally extended to directed graphs [59] and weighted graphs [60, 61].

Degree centrality [51], closeness centrality [51], and betweenness centrality [51] are classical yet widely used measures [47].

Degree centrality estimates the influence of a node on the basis of its degree. The degree centrality of node v is defined as

$$C_d(v) = \frac{|\Gamma(v)|}{N-1}, \tag{1}$$

where N is the number of nodes in the network G and $\Gamma(v)$ is a set of neighbor nodes of v.

Closeness centrality estimates the influence of a node based on the distance between the node and other nodes in a network. The closeness centrality of node v is defined as

$$C_c(v) = \frac{N-1}{\sum_{u \in V} d(v, u)}, \tag{2}$$

where $d(v, u)$ is the shortest path length from node v to u.

Betweenness centrality estimates the influence of a node on the basis of the proportion of shortest paths between all other node pairs passing through the node. The betweenness centrality of node v is defined as

$$C_b(v) = \frac{2}{(N-1)(N-2)} \frac{\sum_{s,t \in V, s \neq t \neq v} g_{st}(v)}{\sum_{s,t \in V, s \neq t} g_{st}}, \tag{3}$$

where g_{st} is the number of shortest paths from node s to t, and $g_{st}(v)$ is the number of shortest paths from node s to t through node v.

Various extensions of betweenness centrality have been proposed, and they can be found in [62]. Representative extensions include random-walk betweenness [63], network flow betweenness [61], and routing betweenness [64]. These measures estimate the influence of a node on the basis of paths other than the shortest paths.

There also exist other types of centrality measures. An excellent review of centrality measures can be found in [65]. Popular measures include Katz centrality [66], eigenvector centrality [67], and eccentricity [68].

PageRank [69] is also widely used for estimating the influence of nodes in social networks, although it was originally proposed for estimating the importance of a web page. The PageRank of node v is defined as

$$PR(v) = \frac{1-\alpha}{N} + \alpha \sum_{u \in \Gamma(v)} \frac{PR(u)}{|\Gamma(u)|}, \tag{4}$$

where α is a damping factor that can be set between 0 and 1. The basic idea of PageRank is that a node linked by influential nodes is influential.

The k**-core** (also called k-shell) index [47, 70–72] estimates the influence of a node according to the size of a dense subnetwork, called the core, to which the node belongs. The k-core of node v in network G is defined as follows. Let H be a subnetwork of network G, and let $\delta(H)$ be the degree of a node whose degree is the minimum among the nodes belonging to subnetwork H. In other words, the degree of each node belonging to H is at least $\delta(H)$. Note that the degree of each node is calculated only from links between

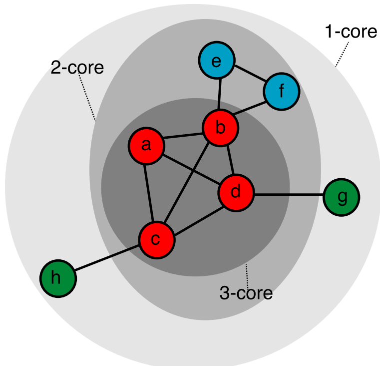


Fig. 3 An example of k-cores and k-core index of nodes. Nodes a, b, c, and d belong to 1-core, 2-core, and 3-core, thus their k-core index is 3. Nodes e and f belong to 1-core, and 2-core, thus their k-core index is 2. Nodes g and h belong to only 1-core, thus their k-core index is 1.

nodes in subnetwork H. Subnetwork H is a k-core of network G if $\delta(H) \geq k$. The k-core index of node v is defined as the maximum k of the k-core to which node v belongs. Figure 3 shows an example of k-cores.

Collective influence (CI) is a recently proposed scalable measure for identifying influential nodes based on the optimal percolation theory [73]. The CI of node v is defined as

$$CI_l(v) = (|\Gamma(v)| - 1) \sum_{u \in \partial Ball(v,l)} (|\Gamma(u)| - 1), \qquad (5)$$

where $\partial Ball(v, l)$ is a set of nodes that are l hops away from node v, with l a parameter. This measure can be used for estimating influence of nodes. Using this measure, Morone and Makse [73] have also proposed an algorithm for finding a set of influential nodes to be removed so that the network connectivity (i.e., the size of the giant component) is minimized. The algorithm is as follows. We first calculate the CI values of all nodes and then obtain the node with the maximum CI value. We then remove the obtained node from the social network G, recalculate the CI values for the remaining nodes, and obtain the node with the maximum CI value. We repeat this procedure and obtain a node ranking based on CI. It is shown that when removing the nodes according to the obtained ranking based on CI, near optimal percolation can be achieved [73]. Moreover, the node ranking is shown to be effective for identifying influential nodes in several social networks [74].

3.3 Influence Maximization Algorithms

As defined in Sect. 3.1, the IM problem aims to find a set of nodes that can influence to many other nodes under a given influence cascade model. We first introduce two representative influence cascade models: the independent cascade (IC) model [17] and the linear threshold (LT) model [17]. We then introduce IM algorithms for these models. Although the IC and LT models are the most popular, IM for other cascade models, such as the credit distribution model [75], the voter model for signed networks [76], and the continuous time diffusion model [77–79], are also being studied. Please refer to the above articles [75, 77–79] for the IM algorithms for other cascade models.

Algorithm 1 Simple greedy IM algorithm [17] (G, σ)

initialize $S \leftarrow \emptyset$
while $|S| < k$ **do**
  select $u \leftarrow \arg\max_{v \in V \setminus S} \sigma(S \cup \{v\}) - \sigma(S)$
  $S \leftarrow S \cup \{u\}$
end while
return S

In the IC model, each node is either active or inactive. When node u becomes active at time step t, node u will influence inactive neighbor node $v ((u, v) \in E)$ with probability $p_{u,v}$ at the next time step $t + 1$. Namely, node v becomes active with probability $p_{u,v}$. The probability $p_{u,v}$ is the weight of link (u, v) and is called the influence spread probability between node u and v. Note that each node has a single chance to influence each of its neighbors. At time step 0, the nodes selected as seed nodes $(S \subseteq V)$ become active, and other nodes are inactive. Then, the stochastic process explained above is repeated until no new active nodes are created, at which time the process ends.

In the LT model, too, each node is either active or inactive. For each node v, the sum of the weights on all incoming links is assumed to be at most 1. Each node v chooses a threshold θ_v uniformly at random from $[0, 1]$. At time step t, each inactive node v becomes active if the sum of the weights on incoming links from v's active neighbors exceeds its threshold θ_v. At time step 0, the nodes selected as seed nodes $(S \subseteq V)$ become active, and other nodes are inactive. Then, the process explained above is repeated until no new active nodes are created, at which time the process ends.

Although the IM problems for both IC and LT models are NP-hard, the objective function $\sigma(\cdot)$ is non-negative, monotone, and submodular, which gives us the greedy $(1 - 1/e)$-approximation algorithm [17]. A set function $f : 2^V \rightarrow \mathbb{R}$ is non-negative if $f(S) \geq 0$ for all $S \subseteq V$, monotone if $f(S) \leq f(T)$ for all $S \subseteq T$, and submodular if $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$ for all $S \subseteq T$ and $v \in V$. The greedy algorithm [17] (simple greedy) starts with an empty seed set $S = \emptyset$ and repeatedly adds the node u with the maximal marginal influence (i.e., $u = \arg\max_{v \in V \setminus S} \sigma(S \cup \{v\}) - \sigma(S)$) into S until $|S| = k$. Pseudocode of Kempe's greedy algorithm is shown as Algorithm 1.

Although the greedy algorithm is very simple and gives us a theoretical guarantee, it requires high computational costs [80, 81]. Since the exact computation of $\sigma(\cdot)$ is #P-hard [80], Monte-Carlo simulation has been used to obtain $\sigma(\cdot)$ [17, 81]. We should note that the simple greedy algorithm provides $(1 - 1/e)$-approximation solution only if the exact influence spread $\sigma(\cdot)$ can be calculated. To obtain an accurate estimation of $\sigma(\cdot)$, we require a large number of simulation runs. The problem of calculating influence

Table 3 Influence measures and influence maximization algorithms.

Measures	Description	Notes	Required info.
Degree [51]	Nodes with high degree are influential	simple yet widely used	local
Closeness [51]	Nodes who have short paths to other nodes are influential		global
Betweenness [51]	Nodes bridging many paths are influential	widely used yet computationally expensive	global
PageRank [69]	Nodes who are linked by influential nodes are influential		global
k-core [47, 70–72]	Nodes in *cores* of the network are influential		global
CI [73]	Nodes whose removal affects the network connectivity are influential	recently proposed and effective	semi-local
IM algorithms			
Simple greedy [17]	Using Monte-Carlo simulation of influence cascades	computationally very expensive	global
TIM [86], IMM [79]	Using sampling technique for estimating influence spread	state-of-the-art	global
IRIE [89]	Heuristics based on the IC model	not using Monte-Carlo simulation	global
SIMPATH [90]	Heuristics based on the LT model	not using Monte-Carlo simulation	global

spread is discussed in [82].

Several improvements in the computational costs of IM algorithms have been proposed [79, 83–88]. Cost-effective lazy forward selection (CELF) [83] exploits the submodularity to reduce the cost of calculating the influence spread, which allows it to be more efficient than the simple greedy algorithm. An improved version, called CELF++ [84], has been also proposed. Recently, Borgs et al. [85] made a breakthrough that achieves near-linear computational time using a sampling technique called reverse influence sampling (RIS). Tang et al. proposed an algorithm, two-phase influence maximization (TIM) [86] and an improvement, influence maximization via martingales (IMM) [79] using the idea in [85]. Further improvements over IMM can be found in [87, 88]. Thanks to the efforts of many researchers, state-of-the-art IM algorithms can work on huge-scale networks with millions of nodes [79, 87, 88].

Along another line, heuristic algorithms that avoid accurate estimation of influence spread have also been studied. Among these, the IRIE (influence ranking and influence estimation) [89] and SIMPATH [90] algorithms are efficient algorithms for the IC model and the LT model, respectively. These algorithms estimate the influence of nodes without Monte-Carlo simulation. Note that influence measures introduced in Sect. 3.2 can be used as heuristic algorithms for the IM problem.

3.4 Related Research Topics

While the measures and algorithms introduced above use only a social network for identifying influencers, several other approaches that use the combination of a social network and other information have been proposed. TwitterRank [91] estimates the influence of Twitter users using social network topology and topics [92] obtained from their tweets. Bakshy et al. [93] identified influencers on Twitter from records of information cascades using a machine-learning technique.

Identifying influential nodes from only limited knowledge on a social network has been also studied. Mihara et al. studied the IM problem for unknown social networks [94, 95]. They proposed IM algorithms that only use partially observed social networks. Kim et al. [96] studied the problem of finding influential neighbors of each node.

Influence measures for temporal networks and multilayer networks have also been proposed. For temporal networks, eigenvector-based centrality [97] and random walk centrality [98] have been proposed, and centrality measures [99] and PageRank [100] have been proposed for multilayer networks.

Table 3 summarizes the representative influence measures and influence maximization algorithms introduced in this section. The different measures and algorithms are based on different ideas. Among the IM algorithms, IMM is state-of-the-art in terms of efficiency and effectiveness. When using IM algorithms shown in this table, careful construction of a social network is necessary because these algorithms require complete global knowledge of the social network, including link weights. It is difficult to determine which influence measures are the best, since the definitions of influence can be different among different applications. Therefore, it is necessary to choose appropriate influence measures for the application, according to their computational costs and the required information.

4. Link Prediction

An important task in SNA related to link-level characteristics is link prediction [29], which predicts future link formation in a given social network. Link prediction has many applications, such as recommendation [18, 29], anomaly detection [101], network modeling [102], missing link detection [103], and the evaluation of network evolution mechanisms [104].

In socially aware networking, link prediction is used for measuring the similarity of nodes. Link prediction techniques are mainly used to identify node pairs that have a high probability of having contact, which helps routing in mobile opportunistic networks [20].

This section first gives a definition of the link prediction problem (Sect. 4.1). We then introduce link prediction techniques that measure the likelihood of link formation (i.e., similarity) between two nodes (Sects. 4.2 and 4.3). Finally, we briefly review research topics related to link prediction (Sect. 4.4).

4.1 Problem Formulation

Typically, link prediction is the problem of either detecting

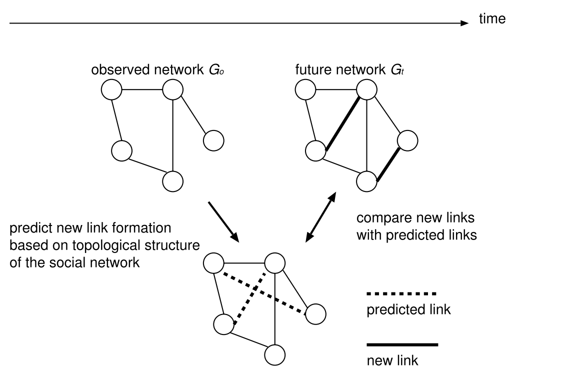


Fig. 4 Illustrative example of future link prediction: New link formation in network G_t is predicted from network G_o.

missing links or predicting future link formation in a network by utilizing a given network topology [29]. Formally, let $G_o = (V, E_o)$ and $G_t = (V, E_t)$ be undirected unweighted networks, where network G_o represents the observed network and network G_t represents the future network or true network for which we would like to predict the links. The link prediction problem is defined as follows, and an illustrative example of future link prediction is shown in Fig. 4.

Problem 3 (Link Prediction Problem) *For each node pair $(i, j) \notin E_o$, the aim is to predict whether $(i, j) \in E_t$, using the observed network G_o.*

Although this paper assumes G_o is an unweighted and undirected network, most link prediction techniques can be naturally extended to weighted networks [105, 106] and directed networks [107, 108].

In what follows, we focus on *unsupervised* link prediction in particular. Unsupervised link prediction techniques estimate the likelihood of link formation (i.e., produce a link prediction score) between two nodes by using knowledge about the characteristics of real networks. For each node pair $(i, j) \notin E_o$, we calculate the link prediction score $s(i, j)$, which estimates the likelihood of link formation or the existence of a link between the nodes. Link prediction score $s(i, j)$ is widely used for measuring the *similarity* of nodes in socially aware networking [20]. Moreover, link prediction scores can be used in supervised link prediction as features for learning [109]. Therefore, we focus on unsupervised link prediction. Link prediction scores used in unsupervised prediction are introduced in Sects. 4.2 and 4.3. Supervised link prediction techniques are briefly introduced in Sect. 4.4.

4.2 Neighborhood-Based Measures

Neighborhood-based link prediction uses only local information. Namely, link formation between node u and v is predicted from information about neighbors of nodes u and v.

The **common neighbors (CN)** method predicts link formation based on the idea that the existence of many common adjacent nodes between two nodes implies a high probability of new link formation between those two nodes [110]. In CN [110], $s(u, v)$, an estimate of the likelihood of link formation between nodes u and v, is given by

$$s_{\mathrm{CN}}(u, v) = |\Gamma(u) \cap \Gamma(v)|, \tag{6}$$

where $\Gamma(u)$ is the set of nodes adjacent to node u.

The **Jaccard coefficient (JC)** method predicts new link formation from the number of common adjacent nodes, similarly to CN, but the link prediction score is normalized [111]. In JC [111], the link prediction score is given by

$$s_{\mathrm{JC}}(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}. \tag{7}$$

The **Adamic/Adar (AA)** method predicts new link formation from the idea that many common adjacent nodes with a small degree between two nodes implies a high probability of new link formation between the nodes [112]. Similarly to CN, AA predicts new link formation on the basis of the number of adjacent nodes in common, but it assigns a weight to $s(u, v)$ according to the degrees of the common adjacent nodes. In AA [112], the link prediction score is given by

$$s_{\mathrm{AA}}(u, v) = \sum_{k \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log |\Gamma(k)|}. \tag{8}$$

The **resource allocation (RA)** method predicts new link formation on the basis of a similar idea as AA [113]. RA assigns a weight to $s(u, v)$ based on the degree of common adjacent nodes. In RA [113], the link prediction score is given by

$$s_{\mathrm{RA}}(u, v) = \sum_{k \in \Gamma(u) \cap \Gamma(v)} \frac{1}{|\Gamma(k)|}. \tag{9}$$

The **preferential attachment (PA)** method predicts new link formation from the idea that a high-degree node has a higher chance of forming new links [110]. In PA [110], the link prediction score is given by

$$s_{\mathrm{PA}}(u, v) = |\Gamma(u)| \times |\Gamma(v)|. \tag{10}$$

4.3 Path-based Measures

In contrast with neighborhood-based link-prediction, path-based link prediction uses global knowledge of the social network G to predict links.

The **Katz index** can be used to predict new link formation based on all the paths between two nodes [66]. The link prediction score of the Katz index is given by

$$s_{\mathrm{Katz}}(u, v) = \sum_{l=1}^{\infty} \beta^l |\mathrm{paths}_{u,v}(l)|, \tag{11}$$

where $\mathrm{paths}_{u,v}(l)$ is a set of all paths with length l between u and v, and β is a parameter. The Katz index gives shorter paths higher weight. Since calculating Katz index for large-scale social networks is computationally expensive, truncated Katz index defined by the following equation is also

Table 4 Link prediction techniques (node similarity scores).

	Description	type
Common neighbor (CN) [110]	Number of common neighbors between two nodes	local
Jaccard coefficient (JC) [111]	Normalized CN	local
Adamic/Adar (AA) [112]	Same idea as CN but low-degree neighbors have larger weight	local
Resource allocation (RA) [113]	Same idea as AA but weight function is different	local
Preferential attachment (PA) [110]	Hub nodes get higher chance to get new links	local
Katz index [66]	Number of paths between two nodes	global
Random walk with restart (RWR) [29]	Nodes who have many short paths between them are similar	global
SimRank [115]	Two nodes are similar if their neighbors are similar	global

used [114].

$$s_{\mathrm{tKatz}}(u,v) = \sum_{l=1}^{l_{max}} \beta^l |\mathrm{paths}_{u,v}(l)|, \tag{12}$$

where l_{max} is a parameter.

The **random walk with restart (RWR)** [29] technique is based on PageRank [69]. Consider a random walker on a social network G from node u, who returns to node u with probability α. Namely, with probability α, a random walker jumps to the starting node u, and with probability $(1-\alpha)$, the random walker moves to the randomly selected neighbor of the current node. Then, the link prediction score between node u and v is defined as the probability that the random walker starting from node u will be located at node v in the steady state.

SimRank [115] is based on the assumption that two nodes are similar if they are connected to similar nodes. The link prediction score of SimRank is given by

$$s_{\mathrm{SR}}(u,v) = \alpha \frac{\sum_{t\in\Gamma(u)} \sum_{t'\in\Gamma(v)} s_{\mathrm{SR}}(t,t')}{|\Gamma(u)||\Gamma(v)|}, \tag{13}$$

where $s_{\mathrm{SR}}(u,u) = 1$ and $\alpha \in [0,1]$ is a parameter. The SimRank can be understood as characterizing a random walk. SimRank measures how soon two random walkers who start from nodes u and v are expected to meet at a certain node.

4.4 Related Research Topics

Supervised approaches for link prediction have also been studied [109, 116–119]. Supervised approaches construct a classifier that can predict whether a link exists between two nodes by using several features obtained from the network topology. Link prediction scores can be used as features for learning. Kashima and Abe [117] proposed a parameterized probabilistic model for link prediction. Lichtenwalter et al. [109] used general classifiers and examined the effectiveness of several topological features. Pujari and Kanawati [120] proposed supervised rank aggregation. Recently, a deep-learning-based approach was also proposed [119]. Other techniques can be found in a recent survey paper [121].

The incorporation of temporal information was shown to be a promising approach [119,122,123] for improving link prediction accuracy. The link prediction scores introduced in this section were obtained from a single snapshot of a social network. However, for the link prediction problem, temporal information (i.e., information about when the existing links were created) is important. Tylenda et al. [122] proposed link prediction scores incorporating temporal information. Tsugawa and Ohsaki [123] examined the construction method of a social network used for link prediction that incorporated temporal information. In the recently proposed deep-learning-based link prediction [119], the observed social network is modeled as a temporal network.

Link prediction in multilayer networks is also a hot research topic [124, 125]. These studies show that using metrics obtained from multiple layers greatly improves the performance of both unsupervised and supervised link prediction compared with using only metrics obtained from a single network.

Table 4 summarizes major link-prediction techniques introduced in this section. The effectiveness of these techniques depends on the network characteristics. Comparative studies of several link prediction techniques can be found in [29, 30].

5. Community Detection

Many social networks have been shown to have *community* structure, where the network is composed of densely connected subgraphs and sparse links connecting the subgraphs [13, 16, 126]. An example of communities in a social network is shown in Fig. 5. The densely connected subgraphs are called communities, and algorithms detecting such communities have been actively studied [32, 33, 126]. Since the communities obtained from social networks are shown to reflect groups of individuals with similar characteristics or backgrounds, community detection algorithms have a wide variety of application domains, such as recommender systems [127], viral marketing [128], and predicting information diffusion [19].

In socially aware networking, communities in a social network are used for measuring the similarity of nodes [20]. Community-based socially aware routing schemes have been proposed, based on the assumption that nodes in the same community will meet more frequently than nodes in different communities [20]. In socially aware caching schemes, communities are used for measuring the similarity of interests between nodes [20].

In what follows, we first give a definition of the community detection problem (Sect. 5.1). We then introduce community detection algorithms (Sects. 5.2 and 5.3). We

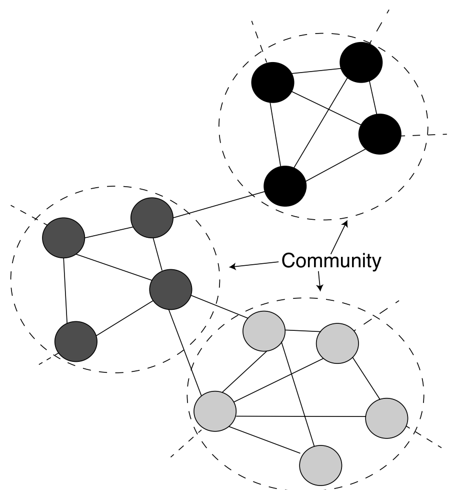


Fig. 5 An example of communities in a social network.

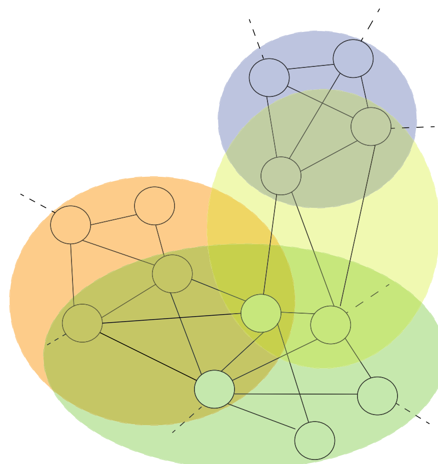


Fig. 6 An example of overlapping communities in a social network.

also discuss other related research topics on community detection in Sect. 5.4.

5.1 Problem Formulation

First of all, it should be noted that there is no universal definition of communities and the community detection problem [32]. Several definitions of communities can be found in the literature [32, 129]. For the variety of definitions and metrics for evaluating community structure, please refer to the excellent survey papers [32, 129]. However, there is one widely accepted basic concept of a community: there must be more links within the community than links connecting to nodes outside the community [32]. Therefore, the community detection problem can be informally considered as a problem of finding such subgraphs in a given network.

Although there are several formulations of the community detection problem, they can be categorized into *disjoint* community detection and *overlapping* community detection. Disjoint community detection can be considered as a graph partitioning [32]. In disjoint community detection, each node belongs to exactly one community (Fig. 5). In contrast, overlapping community detection allows each node to belong to multiple communities (Fig. 6). In our society, it is natural to assume that each individual belongs to multiple communities. Overlapping community detection aims to find such complex structures. More formal definitions of disjoint communities and overlapping communities are given as follows.

Definition 1 (Disjoint Communities) *A set of disjoint communities $C = \{c_1, c_2, \ldots, c_k\}$ in social network $G = (V, E)$ satisfies the following conditions: (1) c_i is a nonempty subset of nodes ($\forall i, V \supseteq c_i \neq \emptyset$), (2) $\bigcup_{i=1}^{k} c_i = V$, (3) for any two nodes $u, v \in c_i$, there exists at least one path where all nodes along the path are also in c_i, and (4) $c_i \cap c_j = \emptyset$ for $i \neq j$.*

Definition 2 (Overlapping Communities) *A set of overlapping communities $C = \{c_1, c_2, \ldots, c_k\}$ in social network $G = (V, E)$ satisfies the following conditions: (1) c_i is a nonempty subset of nodes ($\forall i, V \supseteq c_i \neq \emptyset$), (2) $\bigcup_{i=1}^{k} c_i = V$, (3) for any two nodes $u, v \in c_i$, and there exists at least one path where all nodes along the path are also in c_i.*

Other typologies of community detection also exist, such as *hierarchical* and *non-hierarchical* [32]. In our society, groups are nested, that is, small groups compose larger ones [130]. Detecting such hierarchical community structures has also been studied [32, 130–132].

In this paper, we introduce disjoint community detection algorithms in Sect. 5.2 and overlapping community detection algorithms in Sect. 5.3. We particularly focus on non-hierarchical community detection, but some algorithms introduced in this paper can be used to obtain hierarchical communities, which will be mentioned in the following sections. In what follows, we assume that a social network G is undirected and unweighted. Most algorithms can be easily extended to weighted networks (e.g., see [32, 33]). Defining communities in directed networks is nontrivial, and therefore several studies on community detection algorithms for directed networks have been performed [133, 134]. However, we can apply algorithms for undirected networks by simply ignoring link direction (e.g., [3, 19]).

5.2 Disjoint Community Detection

Among the various approaches of community detection, *modularity* maximization is particularly widely used [32]. In this approach, the community detection problem is formulated as an optimization problem that aims to find a set of communities that maximizes the modularity score [16], which is a measure for evaluating the quality of community detection. The modularity score of the detected disjoint communities C in graph $G = (V, E)$ is defined as

$$Q(C) = \sum_{c_i \in C} \left(\frac{e_{c_i}}{|E|} - \left(\frac{a_{c_i}}{2|E|}\right)^2\right), \tag{14}$$

where e_{c_i} is the number of links connecting nodes in community c_i and a_{c_i} is the sum of the degrees of the nodes in

Algorithm 2 GN algorithm [16, 135] ($G = (V, E)$)

```
1: initialize ComList ← ∅
2: while E ≠ ∅ do
3:     find link e(u, v) with highest betweenness in G
4:     remove link e(u, v) from E
5:     set C as a set of connected components in G
6:     m ← modularity of current communities C
7:     add (C, m) to ComList
8: end while
9: return C with the highest modularity from ComList
```

community c_i. Modularity evaluates the tradeoff between the fraction of intra-community links and the expected fraction of such links when the links are randomly rewired keeping the degree distribution [16]. Using this measure, modularity maximization is defined as follows.

Problem 4 (Modularity Maximization) *Given a social network $G = (V, E)$, the aim is to find a set of communities C such that the modularity score $Q(C)$ is maximized.*

The first algorithm using modularity was proposed by Girvan and Newman [16, 135]. This was a pioneering work of community detection in the network science research field. The basic idea of the Girvan Newman (GN) algorithm is that links connecting communities have high link betweenness [135]. Similarly to the node betweenness [51] introduced in Sect. 3.2, the betweenness of a link is defined as the number of shortest paths passing through the link. Based on this idea, the GN algorithm repeatedly removes the link with the highest betweenness from the network. This procedure causes the network to be fragmented into several components, and then each of the components is considered as a community. The difficulty with using GN is deciding when to stop the link removal process. In the GN algorithm, the modularity score is used as the stopping criteria [16]. Link removal is repeated until all links are removed, and for each removal, the modularity score is calculated. Then, the result with the highest modularity is used. Pseudocode of the GN algorithm is shown as Algorithm 2. Although the GN algorithm uses modularity to determine when to stop, it does not directly maximize the modularity score.

After the GN algorithm was proposed, many algorithms that aim to maximize modularity were proposed. Since modularity maximization problem is NP-hard [136], many greedy heuristic approaches have been proposed for approximation. Newman [137] proposed a greedy algorithm for modularity maximization. The Newman algorithm is an agglomerative hierarchical clustering. Starting from the set of singleton communities (i.e., the set in which each node is considered as a community), the Newman algorithm greedily merges communities by choosing the merge that gives the highest increase in modularity value. Clauset et al. further optimized the Newman algorithm, proposing a faster algorithm called the CNM algorithm [138]. The Louvain algorithm [139] uses local optimization and achieves lower computational costs than the CNM algorithm. It was also shown that the Louvain algorithm generally achieves a higher modularity score than CNM [139]. The Louvain algorithm is widely used and is one of the state-of-the-art algorithms based on modularity. Further improvement and extensions of the Louvain algorithm can be found in [140]. Other approaches to maximizing modularity include simulated annealing [141], spectral optimization [142], and external optimization [143].

Although modularity-based clustering (e.g., CNM and Louvain algorithms) has been widely applied to many applications (e.g., [3, 144, 145]), it is important to be aware of the limitations of modularity-based algorithms [32, 33, 146] when using them. The first limitation is that a high modularity score can be achieved even if the target network has no community structure [147]. Even for a random network, modularity maximization may discover communities with high modularity score. Another limitation is the resolution limit of modularity [148]. Modularity maximization algorithms may fail to detect small communities in a network due to this limitation [148]. Moreover, it has been shown that there are several different community partitions in a network whose modularity scores are very close to the global maximum value [149]. Although greedy algorithms and the GN algorithm can produce hierarchical community structures, it has been suggested that it is unclear whether some of the intermediate partitions could correspond to meaningful hierarchical levels of the graph [32]. Due to the limitations mentioned above, it is important to take care with the results when using the modularity-based algorithms. However, in our opinion, modularity-based community detection is still a useful tool if the target social network has a clear community structure.

5.3 Overlapping Community Detection

Overlapping community detection is a relatively new and challenging research topic compared with disjoint community detection [32–34].

The clique percolation method (CPM), proposed by Palla et al. [150], is one popular algorithm. A k-clique is a complete subgraph with k nodes. CPM finds overlapping communities by finding k-cliques. In CPM, two k-cliques are considered to be adjacent if they share $(k - 1)$ nodes. Then, a k-clique community is defined as the union of all k-cliques that can be reached from each other through a series of adjacent k-cliques. Each node may belong to multiple k-clique communities. Thus, CPM can produce overlapping communities. Kumpula et al. [151] proposed faster algorithms for CPM.

Another popular approach is link clustering. Ahn et al. [152] proposed measuring the similarity of two links (u, k) and (v, k) with the Jaccard coefficient between node u and v defined as Eq. (7). Using this similarity measure, hierarchical clustering of links is performed. Starting from singleton communities, similar links are repeatedly merged into a community. Link communities can be obtained by stopping the merging process at a similarity threshold. Since each node can belong to multiple link communities, overlap-

Table 5 Community detection algorithms.

	Notes	overlapping	hierarchical
GN [16, 135]	Pioneering algorithm. Computationally expensive.		✓
CNM [138]	Modularity maximization. Faster than GN.		✓
Louvain [139]	Faster than CNM and achieves higher modularity than CNM		✓
CPM [150]	Based on clique percolation	✓	
Link community [152]	Clustering links rather than nodes	✓	✓
LFM [130]	Using seed set expansion from random nodes	✓	✓
NICE [154]	Using seed set expansion. Seeds are determined based on personalized PageRank clustering.	✓	✓

ping communities can be obtained. An algorithm using a line graph, where links in the original graph are nodes and nodes in the original graph are links, has also been proposed [153]. Evans and Lambiotte [153] converted an original graph to a line graph and applied the disjoint community detection algorithm to the line graph in order to obtain link communities.

Local expansion is an approach for overlapping community detection that is currently being actively studied [154]. In this approach, several seed sets are selected, and each seed set locally expands its community [34]. Namely, each seed set corresponds to a community, and the seed set locally searches for members that should belong to the community. Several strategies for seed node selection and expansion exist. Lancichinetti et al. [130] proposed an algorithm called LFM that expands a community from a random seed node to form a *natural* community. The *naturalness* of community c_i is evaluated by a fitness function defined as

$$f(c_i) = \frac{k_{\text{in}}(c_i)}{(k_{\text{in}}(c_i) + k_{\text{out}}(c_i))^{\alpha}}, \tag{15}$$

where $k_{\text{in}}(c_i)$ and $k_{\text{out}}(c_i)$ are the total internal and external degrees of community c_i, respectively, and α is a parameter for controlling the community size. The proposed method expands a community until this fitness function is locally maximal. Havemann et al. [155] modified the fitness function and proposed a parameter-free algorithm. An algorithm called EAGLE [131] uses maximal cliques as seed sets. Whang et al. [154] proposed an algorithm called NICE to use high-degree nodes as seed sets and expand the seed sets using personalized PageRank clustering [156]. These algorithms also explicitly consider hierarchical community structures.

5.4 Related Research Topics

Many approaches for community detection exist. We will briefly introduce other popular approaches. Optimization-based approaches using metrics other than modularity have been studied, with examples being optimization based on normalized cuts [157] and conductance [158]. Non-negative matrix factorization [159] and random walk [132] have also been used for overlapping and hierarchical community detection. A recent trend in community detection is using statistical inference [33, 126, 160–162]. In this approach, a network is assumed to be generated by a probabilistic model that has a community structure, and the goal is to find the model that best fits the observed network. A survey on this approach has recently become available in [162].

As we have already discussed, many community detection algorithms have been proposed. Therefore, comparison of these algorithms has been performed [34, 163–165]. However, to the best of our knowledge, there is no consensus on which algorithm is the best [33]. This is because community can be defined in many ways, meaning that the best algorithm depends on the applications of the community detection. When choosing the appropriate algorithm for your problem, it is important to know the characteristics of the algorithms (e.g., resolution limits of modularity-based algorithms), and the above studies provide useful information.

Community detection in multilayer networks and temporal networks are also topics of high interest. For community detection for temporal networks (or dynamic networks) please refer to [166–168], and for multilayer networks please refer to [125, 166].

Table 5 summarizes the major community detection algorithms introduced in this section.

6. Socially Aware Networking

This section first introduces the research background of socially aware networking (Sect. 6.1) and then explains how SNA can be used for designing socially aware networking schemes (Sects. 6.2 and 6.3).

6.1 Background

Designing networking technologies that take human behaviors into consideration is a topic that has been getting increased attention from many researchers, with the aim of achieving efficient utilization of limited network resources and improved quality of service by various applications [4, 20]. Particularly for mobile networks, usage patterns and mobility patterns of network users are suggested to significantly affect the performance of communication networks [20]. Thus, incorporating human factors is a promising candidate for future networking technologies.

Socially aware networking is an emerging research field that aims to improve the current network technologies and realize novel network services by utilizing social relationships among network users [20]. Socially aware networking technologies are currently expected to be applied to mobile networks, such as mobile opportunistic networks [169], delay/disruption tolerant networks (DTNs) [170], and vehicular ad-hoc networks (VANETs) [171]. In such networks, messages (i.e., data or packets) are transmitted in a multi-hop manner among mobile devices with short-range wireless communication, meaning that the mobility patterns of devices strongly affect the efficiency of message delivery. Since the mobile devices are mainly carried by individuals, the mobility patterns are strongly affected by the characteristics of individuals and the social relationships among them [20]. Thus, social networks are expected to be a useful source of information that can be used for improving the quality of services in such mobile networks.

Two of the main research topics in the socially aware networking research field are *socially aware routing* and *socially aware caching*. In what follows, we introduce studies on both technologies. Note that game-theoretic approaches are out of scope of this paper, although they also incorporate human factors in the design and control of communication networks. This paper particularly focuses on the studies using SNA techniques. Game-theoretic approaches can be found in papers such as [172–175].

6.2 Socially Aware Routing

6.2.1 Overview

One of the most active topics in the socially aware networking research field is routing/forwarding schemes in opportunistic networks [20]. Opportunistic networks, or DTNs, are attracting attention from many researchers as promising platforms for realizing communication without constructing additional infrastructure [169, 176]. Many socially aware routing schemes for such networks have been proposed [20, 177–191]. Here, we consider the problem of delivering message M from node s to node t in an opportunistic network. In an opportunistic network, the existence of a path between the source and destination nodes cannot be assumed [169]. Therefore, many opportunistic network routing schemes adopt the store-carry-and-forward paradigm for message delivery [176]. In such schemes, when node u with message M meets node v (i.e., node u and v are close to each other), message M can be forwarded to node v. By repeating such message forwarding, a message will be delivered from source node s to destination node t. The simplest way to achieve message delivery in opportunistic networks is by using epidemic routing [192], in which each node forwards messages to every encountered node. While epidemic routing achieves optimally low message delivery delay, it consumes significant network resources. Therefore, forwarding a message to relay nodes that have a high probability of meeting the destination node is a key issue in the design of routing schemes [176]. Socially aware routing schemes utilize SNA for estimating the likelihoods of future contacts among nodes, and the estimated likelihoods are used in message forwarding [20].

An illustrative explanation of socially aware routing scheme is shown in Fig. 7. In socially aware routing, contact probabilities between two individuals are estimated from influence, similarity (i.e., link prediction scores), and/or community. An intuitive explanation of key ideas in socially aware routing is as follows. (1) Influential nodes should deliver messages to many other nodes that are suitable as relay nodes. (2) Similar nodes will have a high probability of contact. (3) Nodes in the same community will have a high probability of contact.

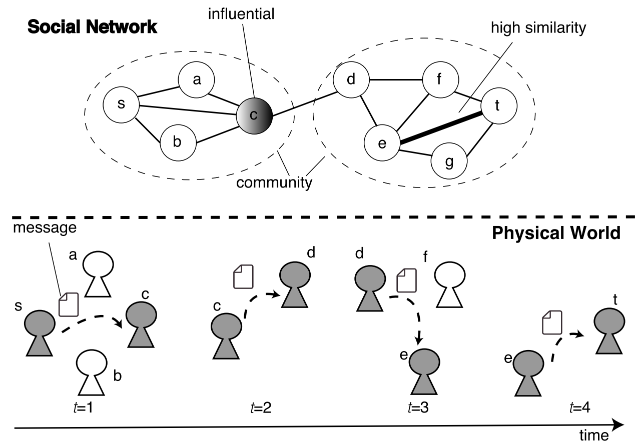


Fig. 7 An illustrative example of socially aware routing: Suppose that node s wants to send a message to node t. (t = 1): Node s forwards the message to node c since node c is influential and is expected to meet many other nodes. (t = 2): Node c forwards the massage to node d since node d and the destination node t are in the same community, which implies that they will have contacts with high probability. (t = 3): Node d forwards the message to node e since node e and the destination node t are similar, which implies that they will have contacts with high probability. (t = 4): Finally, node e meets node t, and the message is successfully delivered to the destination node t.

Before introducing socially aware routing schemes, we should mention the problem of how to obtain social networks. Obtaining social networks among users is a necessary step shared in common among all socially aware routing schemes [178]. If a social network is available from SNS or questionnaire data, that network can be used in socially aware networking schemes [177]. However, such information is currently difficult to obtain. Therefore, many routing schemes construct social networks from past contact logs, called contact graphs [178]. A contact graph is typically constructed from records of contacts within a specific time window [178]. However, Hossmann et al. [193, 194] showed that the method for constructing a social network (here, a contact graph) significantly affects the performance of socially aware routing. Therefore, the method for constructing social graphs has been studied [195]. Many routing schemes also assume that global knowledge of the social network is unavailable, and the social network is only constructed via distributed information from the past contact logs of each node [177].

Table 6 summarizes major socially aware routing schemes. Different schemes use different social features (i.e., influence, similarity, and community). In what follows, we briefly introduce each scheme and discuss how SNA techniques are used in the schemes.

Table 6 Socially aware routing schemes.

	Social features			
	Community	Influence	Similarity	Source of Social Network
LABEL [196]	✓			Questionnaire
BUBBLE Rap [179]	✓	✓		Contact Logs
LocalCom [183]	✓	✓	✓	Contact Logs
mGroup [185]	✓			Contact Logs, Questionnaire
CAOR [186]	✓	✓		Contact Logs
SMART [187]	✓	✓	✓	Contact Logs
FBR [188]	✓		✓	Contact Logs
dLifecomm [197]	✓	✓	✓	Contact Logs
SimBet [180], SimBetTs [181]		✓	✓	Contact Logs
SRBet [190]		✓	✓	Contact logs
PeopleRank [182]		✓		Contact logs, SNS
SANE [198]			✓	Questionnaire
ML-SOR [184, 189]		✓	✓	Contact Logs, SNS, Questionnaire
dLife [197]		✓	✓	Contact Logs

6.2.2 Community-Based Schemes

Although different schemes use different combinations of social features, as shown in Table 6, socially aware routing can be divided into community-based and non-community-based [20]. The basic assumption in community-based routing is that people in the same community meet each other more frequently than people in different communities. This subsection introduces major community-based schemes.

Hui and Crowcroft [196] performed pioneering work on community-based socially aware routing, proposing a routing scheme called LABEL. LABEL assumes that each node has a label representing its community. In LABEL, node v forwards a message to node u if and only if node u belongs to the same community as the destination node of the message. LABEL was shown to successfully reduce the message overhead while maintaining the message delivery delay using this small amount of information (i.e., community label) [196]. However, the performance of LABEL is heavily dependent on the community, and it has been suggested that, for instance, message delivery will fail when the source node does not meet any node in the destination's community [178].

Hui et al. [179] proposed a routing scheme called BUBBLE Rap, which utilizes influence measures (i.e., degree and betweenness centrality) as well as community information of nodes. BUBBLE Rap uses two metrics: global influence rank, which is a ranking based on an influence measure in a whole social network, and local influence rank, which is a ranking based on an influence measure within a community only. The global influence rank is used for inter-community routing, whereas the local influence rank is used for intra-community routing. A message is forwarded to nodes with higher global influence rank until the message reaches a node that belongs to the same community as the destination node. When a message reaches a destination community, the message is forwarded to nodes with higher local influence rank. Since global knowledge of a social network is difficult to obtain, BUBBLE Rap adopts algorithms for detecting communities and calculating centrality from only local information of each node.

Several other schemes for community-based routing exist. In LocalCom [183], the source node forwards a message to the bridge node, which has high betweenness centrality in the community of the source node, and then the bridge node performs controlled flooding to deliver the message to the destination community. Communities are detected in a distributed way using similarity among nodes. SMART [187] uses a strategy similar to that of BUBBLE Rap. It combines influence and similarity measures for the routing metrics. CAOR [186] assumes that users have frequently visited home locations. From this, home-aware communities are defined, and intra-community and inter-community routing utilizing influence measures is proposed. FBR [188] defines a similarity measure called the social pressure metric (SPM), and friendship communities are detected on the basis of SPM. The forwarding strategy is similar to LABEL, but the friendship communities are periodically updated. mGroup [185] uses multiple communities of each node obtained from different social networks (e.g., social networks obtained from contact logs and SNS).

In summary, community-based routing utilizes the characteristic that two individuals in the same community have a high probability of contact for message routing. Influence and similarity measures are also used to achieve efficient intra-community and inter-community routing. Community detection from only locally available information is key for achieving efficient community-based routing [20]. Therefore, distributed community detection that can be used for community-based routing has also been an important research topic [199, 200].

6.2.3 Non-Community-Based Schemes

To avoid the difficulty of detecting communities, non-community-based approaches have been also studied. These approaches utilize a similarity score between two nodes or an influence score of each node to determine the relay node. This subsection introduces major non-community-based schemes.

SimBet [180] and its extended version SimBetTs [181] use similarity and influence. CN [110] is used as the similarity measure and betweenness is used as the influence measure. Although betweenness centrality is a global measure that requires an entire network, SimBet uses each node's ego-betweenness [201], which can be calculated from a subnetwork including the node and its one-hop neighbors. Since ego-betweenness has a high correlation with global betweenness [201], such an approximation is used. SimBet determines the relay node based on the metric combining the betweenness of the node and the similarity between the relay node and the destination node.

PeopleRank [182] forwards a message based on the influence of a relay node. The influence is measured by the measure PeopleRank, which is based on PageRank. SANE [198] forwards messages based on the similarity between nodes. Similarity between nodes is obtained from the predefined social profile of each node.

Since temporal patterns of human-to-human contacts affect the performance of routing schemes, those utilizing temporal social networks have been proposed. dLife is a pioneering scheme that utilizes a temporal social network for routing in opportunistic networks [197]. The forwarding strategy of dLife is similar to other routing schemes introduced above. If node u encounters node v and node v has higher similarity with the destination node than node u, node u forwards the message to node v. If the similarity is unknown, node u forwards the message to node v if node v has higher influence than node u. The key idea of dLife is estimating similarity based on the daily routines of users. The assumption of dLife is that two individuals who routinely encounter with each other should have high similarity (e.g., if two individuals meet each other at 5 p.m. everyday, they should have high similarity). To find such routine, a temporal network constricted from contact logs is used. Note that community-based version of dLife, which is called dLifecomm, has been also proposed [197]. SRBet [190] also utilizes a temporal network. Its routing strategy is based on a similar idea to SimBet, but it uses a temporal network for calculating similarity and influence. SRBet incorporates contact frequency, contact duration, and regularity of contacts between nodes for estimating the similarity between them. It has been shown that dLife and SRBet achieve better routing efficiency than other socially-aware routing strategies not using temporal networks [190, 197].

Routing schemes using multilayer social networks have been also proposed for reliable calculations of influence and similarity measures [189]. ML-SOR [189] utilizes multilayer social networks for obtaining influence and similarity measures. A multilayer social network is constructed from online social networks and physical contact networks.

In summary, non-community based schemes focus on designing good metrics for relay node selection by combining multiple features obtained from social networks.

6.3 Socially Aware Caching

6.3.1 Overview

Another topic of interest in socially aware networking is content caching. Caching has been getting increased attention as a technology to reduce network traffic and realize efficient content delivery [202, 203].

Content-caching technologies can be categorized into those that cache in the end device and those that cache in the network. Caching in the network is key in content delivery networks (CDNs) [204] and content centric networking (CCN) or named data networking (NDN) [205], and it has been widely studied [202, 203, 206]. More recently, caching in mobile edges (i.e., small-base stations), also called femtocaching [207], has been getting attention as a key technology in 5G wireless networks [203, 208–210]. This is also considered a type of caching in the network. In contrast, the recent development of highly functional mobile devices makes it possible to cache content in the end devices (i.e., mobile devices of network users) [203]. In the future mobile Internet, users are expected to be able to retrieve content not only from the original content server but also from caches located at the neighboring mobile edge devices and network routers [203], reducing content access delay and the network traffic volume.

Socially aware networking paradigms can be applied mainly to caching in the end devices. Since the effectiveness of end device caching heavily depends on human mobility patterns, the social network is useful for the effectiveness of the cache. In socially aware caching, the social network is used for determining where to cache and what to cache. An illustrative explanation of socially aware caching is shown in Fig. 8. The basic assumptions in socially aware content caching are (1) that the proximity between two nodes in a social network should reflect the similarity of interests between the two nodes and (2) that influential nodes are suitable for allocating caches because they have the potential to disseminate the content to many other nodes. We introduce socially aware caching in the end devices in Sect. 6.3.2. Socially aware caching in networks will be briefly introduced in Sect. 6.3.3.

6.3.2 Caching in End Devices

Influence measures and communities are used for determining which nodes should have caches in socially aware caching. Since the caches in the end devices can be accessed by other devices with D2D wireless links, influential nodes who meet many other nodes and who have the potential to disseminate caches are suitable for having caches. Gao et al. proposed allocating caches of popular content to nodes with high centrality [211]. Zhuo et al. [212] proposed a centrality measure for measuring caching capability within a community. Then, they proposed a method to allocate caches to the nodes with high centrality for each community. Bastug et

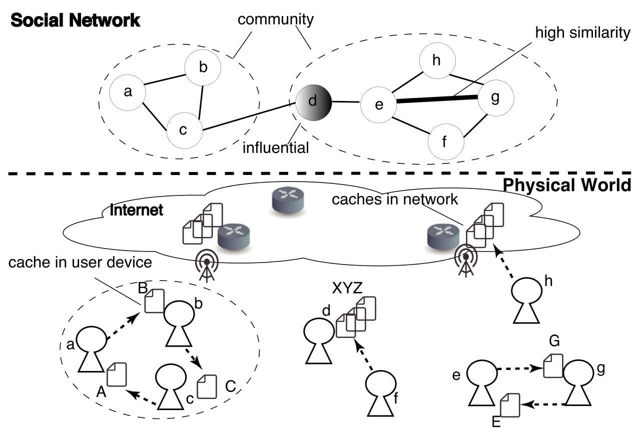


Fig. 8 An illustrative example of socially aware caching: Content caches are located on user devices and network routers. Users can get contents nearby user devices or network routers. Since nodes a, b, and c are in the same community and expected to have similar interest, they share the caches of similar contents A, B, and C. They can access the content caches through D2D links of their devices, which realizes faster access to content than accessing to content repositories (i.e., servers). Similar nodes e and g also share the caches of similar contents. Influential node d has caches of popular contents X, Y, and Z, and those can be accessed by other nodes with D2D links, which contributes to reduce network traffic in the core network.

al [210], and Le et al [213] also proposed similar approaches of cache allocation.

While the above studies aim to allocate caches to a small number of nodes and efficiently disseminate the caches to the other nodes, there are also other approaches where all nodes cooperatively have content caches. In such schemes, it is important to determine what to cache rather than where to cache. To improve the effectiveness of caching, nodes who meet frequently and who have similar interests should have caches of different content. Based on this idea, community-based caching was proposed in [214]. In this scheme, different content caches are allocated for each node in the same community. Namely, each content item is cached in at most one node within a community. Nodes in the same community are expected to meet frequently and have similar interests, so this scheme achieves efficient content caching. Zhu et al. [215] tackled the problem of what to cache by using the similarity of nodes as well as considering the selfishness of nodes.

6.3.3 Caching in Networks

Bernardini et al. [216] studied socially aware caching for CCN and proposed a caching strategy called SACS (Socially aware Caching Strategy). SACS first identifies influencers in social networks. Then, the caches of content produced by the influencers are proactively replicated to the routers along the shortest paths between the influencers and the neighbors of the influencers in the social network.

Wang et al. [217] proposed socially aware content caching (called replication in [217]) in CDN. They used records of video content propagation in social networks for determining which content should be cached at which CDN node. Hu et al. [218] proposed a method for content caching in CDN using social network communities. The basic idea is that each community corresponds to one or several CDN nodes, and all the requests from a community will be served by the corresponding CDN nodes.

6.4 Related Research Topics

In both socially aware routing and caching, the *selfishness* of the nodes is an important issue [20, 203]. The socially aware networking schemes introduced in this paper assume cooperative behavior from each node. Namely, each node is assumed to store and forward messages and cache content for other nodes. However, in reality, it is natural to assume that selfish nodes exist. Incentive mechanisms have been used for mitigating selfish behaviors of individuals [219–221]. Routing schemes incorporating incentive mechanisms can be found in [222–224], and caching schemes incorporating incentive mechanisms can be found in [215, 225].

Security is another important issue [226]. Defense schemes against Sybil attacks [227] and denial of service attacks [228] have been studied in order to prevent attacks from malicious nodes. However, security is still an open issue that should be resolved in order to implement socially aware networking schemes in our society [226].

7. Discussion

7.1 Promising Recent SNA Techniques that can be Applied to Socially Aware Networking

While degree and betweenness centrality are widely used in socially aware networking, as described in Sect. 6, other influence measures are also applicable to socially aware networking technologies. CI [73] is a recently proposed, effective measure for identifying influencers, and it has been shown to be more effective than several other measures, including PageRank, degree, and k-core index in social media networks such as Twitter and Facebook [74]. Moreover, CI can be calculated from semi-local information of each node (i.e., information about l-hop neighbors), which is a preferable feature in socially aware networking. We therefore expect that CI will also be effective in socially aware networking. Influence maximization with limited knowledge on social networks [94, 96] also has the potential to be applied to socially aware networking schemes.

We expect that similarity measures used in socially aware networking can be improved by using perspectives in link prediction studies. While socially aware networking schemes use similarity measures for predicting future contact of individuals, such a problem is also studied in the SNA research field. Scholtz et al. [229–231] and Tsugawa and Ohsaki [123] studied the problem of predicting future human-to-human contact. Takaguchi et al. [232] proposed a method to quantify the predictability of face-to-face contact patterns. We expect that these studies can be applicable to socially aware routing schemes for determining relaying nodes.

We also anticipate that community detection algorithms using local expansion [34, 130, 154, 155] can be applicable to socially aware networking schemes. Local expansion algorithms typically detect communities from only the local network structure of seed nodes [130, 155]. Therefore, these algorithms can be extended to distributed algorithms. Moreover, measures for evaluating local community structure have also been proposed [164, 233]. These can be used to locally evaluate the goodness of a community. While several distributed community detection algorithms have been proposed in community-based socially aware routing studies [179, 188], applying algorithms proposed in SNA studies could be an effective approach.

Moreover, SNA techniques for temporal networks [43, 44] and multilayer networks [41, 42] could be useful in socially aware networking. Contact graphs used in socially aware routing contain information about contact timing, which is suitable for representation as temporal networks. If multiple social networks are available from multiple sources (e.g., SNS data and contact logs), these networks are suitable to be represented as a multilayer network. Using temporal networks, as in dLife [197] and SRBet [190], and using multilayer networks, as in the socially aware routing scheme ML-SOR [189], are expected to be promising approaches.

7.2 Open Issues in Socially Aware Networking

As already discussed in the previous survey papers [20, 177, 226], there are several open issues in socially aware networking, since socially aware networking is an emerging research field. Open issues that should be resolved include privacy, security, selfishness, and scalability.

Of the several open issues, we feel that designing schemes for obtaining social network data in particular is an important issue. As discussed in [226], the amount of available social features affects the performance of socially aware routing schemes. In order to make full use of the benefit of using social networks, only using the past contact logs of each node is not enough, and rich social features are necessary. However, obtaining the social networks of many network users is not easy due to the aforementioned privacy and security issues. Moreover, it is also important to confirm the effectiveness of socially aware networking schemes under ideal situations where rich social features of network users are available.

8. Conclusion

In this paper, we introduced SNA techniques for the identification of influential nodes, link prediction, and community detection, all of which are key techniques for socially aware networking. Then, we discussed how SNA techniques are used in socially aware networking as well as introducing its research trends. Although the research field of socially aware networking is still in its infancy, we believe that socially aware networking has the potential to be a key technology for future communication networks.

Before concluding this paper, we would also like to introduce several tools and datasets useful for SNA and socially aware networking studies. NetworkX[†] and igraph[††] are popular tools for network analysis. Many algorithms introduced in this paper can be easily used with these tools. Gephi [234] and Cytoscape [235] are useful for network visualization. The Stanford Network Analysis Project (SNAP) [236] provides various social network datasets and libraries for SNA. The SocioPatterns project[†††] also provides useful data on human contact. Mobility trace data that is widely used in socially aware networking studies is available through the MIT Reality Mining project [237] and CRAWDAD [238].

We hope that this survey will be helpful to researchers who are interested in using SNA techniques for various problems and will contribute to the development of socially aware networking.

Acknowledgments

This work was partly supported by JSPS KAKENHI Grant Number 16K20931 and 17H01733.

References

[1] D.J. Watts, "A twenty-first century science," Nature, vol.445, no.7127, pp.489–489, 2007. doi:10.1038/445489a.
[2] S.P. Borgatti, A. Mehra, D.J. Brass, and G. Labianca, "Network analysis in the social sciences," Science, vol.323, no.5916, pp.892–895, 2009. doi:10.1126/science.1165821.
[3] S. Tsugawa and H. Ohsaki, "Community structure and interaction locality in social networks," Journal of Information Processing, vol.23, no.4, pp.402–410, 2015. doi:10.2197/ipsjjip.23.402.
[4] S. Niida, S. Tsugawa, M. Suganuma, and N. Wakamiya, "A survey on modeling of human states in communication behavior," IEICE Trans. Commun., vol.100, no.9, pp.1538–1546, 2017. doi:10.1587/transcom.2016PFI0010.
[5] J. Scott, Social Network Analysis, 4th Ed., Sage Publications, 2017.
[6] J. Scott, "Social network analysis: Developments, advances, and prospects," Social Network Analysis and Mining, vol.1, no.1, pp.21–26, 2011. doi:10.1007/s13278-010-0012-6.
[7] N. Eagle, A. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," Proc. National Academy of Sciences, vol.106, no.36, pp.15274–15278, 2009. doi:10.1073/pnas.0900282106.
[8] J.P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.L. Barabási, "Structure and tie strengths in mobile communication networks," Proc. National Academy of Sciences, vol.104, no.18, pp.7332–7336, 2007. doi:10.1073/pnas.0610245104.
[9] H. Koto, H. Nakamura, and M. Aida, "Two-layered structure of social network revealed by data analysis of telecommunciation services," Proc. IEEE Third International Conf. on Social Computing (SocialCom'11), pp.471–476, 2011. doi:10.1109/PASSAT/SocialCom.2011.48.
[10] E. Holger, M. Lutz-Ingo, and B. Stefan, "Scale-free topology of e-mail networks," Phys. Rev. E, vol.66, no.035103, pp.1–4, 2002. doi:10.1103/PhysRevE.66.035103.

† https://networkx.github.io/
†† http://igraph.org/
††† http://www.sociopatterns.org/

[11] G. Kossinets and D.J. Watts, "Empirical analysis of an evolving social network," Science, vol.311, no.5757, pp.88–90, 2006. doi:10.1126/science.1116869.
[12] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," Proc. 19th International Conf. on World Wide Web (WWW'10), pp.591–600, 2010. doi:10.1145/1772690.1772751.
[13] E. Ferrara, "A large-scale community structure analysis in Facebook," EPJ Data Science, vol.1, no.1, 2012. doi:10.1140/epjds9.
[14] B. Viswanath, A. Mislove, M. Cha, and K.P. Gummadi, "On the evolution of user interaction in Facebook," Proc. 2nd ACM SIGCOMM Workshop on Soc. Networks (WOSN'09), pp.37–42, 2009. doi:10.1145/1592665.1592675.
[15] D.J. Watts and S.H. Strogatz, "Collective dynamics of 'small-world' networks," Nature, vol.393, no.6684, pp.440–442, 1998. doi:10.1038/30918.
[16] M.E.J. Newman and M. Girvan, "Finding and evaluating community structure in networks," Phys. Rev. E, vol.69, no.2, p.026113, 2004. doi:10.1103/PhysRevE.69.026113.
[17] D. Kempe, J.M. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," Proc. 9th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD'03), pp.137–146, 2003. doi:10.1145/956750.956769.
[18] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, "WTF: The who to follow service at twitter," Proc. 22nd International Conf. on World Wide Web (WWW'13), pp.505–514, 2013. doi:10.1145/2488388.2488433.
[19] L. Weng, F. Menczer, and Y.Y. Ahn, "Virality prediction and community structure in social networks," Scientific Reports, vol.3, p.2522, 2013. doi:10.1038/srep02522.
[20] F. Xia, L. Liu, J. Li, J. Ma, and A.V. Vasilakos, "Socially aware networking: A survey," IEEE Syst. J., vol.9, no.3, pp.904–921, 2015. doi:10.1109/JSYST.2013.2281262.
[21] M.E.J. Newman, Networks: An Introduction, Oxford University Press, 2010. doi:10.1093/acprof:oso/9780199206650.001.0001.
[22] A.L. Barabási, Network Science, Cambridge University Press, 2016.
[23] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.U. Hwang, "Complex networks: Structure and dynamics," Physics Reports, vol.424, no.4, pp.175–308, 2006. doi:10.1016/j.physrep.2005.10.009.
[24] A. Guille, H. Hacid, C. Favre, and D.A. Zighed, "Information diffusion in online social networks: A survey," ACM Sigmod Record, vol.42, no.2, pp.17–28, 2013. doi:10.1145/2503792.2503797.
[25] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," Wirel. Commun. Mob. Comput., vol.2, no.5, pp.483–502, 2002. doi:10.1002/wcm.72.
[26] L. Lü, D. Chen, X.L. Ren, Q.M. Zhang, Y.C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," Physics Reports, vol.650, pp.1–63, 2016. doi:10.1016/j.physrep.2016.06.007.
[27] S. Bhagat, G. Cormode, and S. Muthukrishnan, "Node classification in social networks," in Social Network Data Analytics, pp.115–148, Springer, 2011. doi:10.1007/978-1-4419-8462-3_5.
[28] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," arXiv preprint, arXiv:1705.02801, 2017.
[29] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," J. Am. Soc. Inform. Sci. Technol., vol.58, no.7, pp.1019–1031, 2007. doi:10.1002/asi.20591.
[30] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," Physica A: Statistical Mechanics and its Applications, vol.390, no.6, pp.1150–1170, 2011. doi:10.1016/j.physa.2010.11.027.
[31] W. Sherchan, S. Nepal, and C. Paris, "A survey of trust in social networks," ACM Comput. Surv., vol.45, no.4, p.47, 2013. doi:10.1145/2501654.2501661.
[32] S. Fortunato, "Community detection in graphs," Physics Reports, vol.486, no.3, pp.75–174, 2010. doi:10.1016/j.physrep.2009.11.002.
[33] S. Fortunato and D. Hric, "Community detection in networks: A user guide," Physics Reports, vol.659, no.11, pp.1–44, 2010. doi:10.1016/j.physrep.2016.09.002.
[34] J. Xie, S. Kelley, and B.K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study," ACM Comput. Surv., vol.45, no.4, pp.43:1–43:35, 2013. doi:10.1145/2501654.2501657.
[35] U. Alon, "Network motifs: Theory and experimental approaches," Nature Reviews Genetics, vol.8, no.6, pp.450–461, 2007. doi:10.1038/nrg2102.
[36] L.d.F. Costa, F.A. Rodrigues, G. Travieso, and P.R. Villas Boas, "Characterization of complex networks: A survey of measurements," Advances in Physics, vol.56, no.1, pp.167–242, 2007. doi:10.1080/00018730601170527.
[37] T. Valente, S. Watkins, M. Jato, A.V.D. Straten, and L. Tsitsol, "Social network associations with contraceptive use among Cameroonian women in voluntary associations," Social Science & Medicine, vol.45, no.5, pp.677–687, 1997. doi:10.1016/S0277-9536(96)00385-1.
[38] M.E.J. Newman, "Coauthorship networks and patterns of scientific collaboration," Proc. National Academy of Sciences, vol.101, no.suppl 1, pp.5200–5205, 2004. doi:10.1073/pnas.0307545100.
[39] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.F. Pinton, and W.V. den Broeck, "What's in a crowd? Analysis of face-to-face behavioral networks," J. Theoretical Biology, vol.271, no.1, pp.166–180, 2011. doi:10.1016/j.jtbi.2010.11.033.
[40] M.S. Granovetter, "The strength of weak ties," Am. J. Sociol., vol.78, no.6, pp.1360–1380, 1973. doi:10.1016/B978-0-12-442450-0.50025-0.
[41] M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, and M.A. Porter, "Multilayer networks," J. Complex Networks, vol.2, no.3, pp.203–271, 2014. doi:10.1093/comnet/cnu016.
[42] S. Boccaletti, G. Bianconi, R. Criado, C.I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," Physics Reports, vol.544, no.1, pp.1–122, 2014. doi:10.1016/j.physrep.2014.07.001.
[43] P. Holme, "Modern temporal network theory: A colloquium," The European Physical Journal B, vol.88, no.9, pp.234:1–234:30, 2015. doi:10.1140/epjb/e2015-60657-4.
[44] P. Holme and J. Saramäki, "Temporal networks," Physics Reports, vol.519, no.3, pp.97–125, 2012. doi:10.1016/j.physrep.2012.03.001.
[45] N. Masuda and R. Lambiotte, A Guide to Temporal Networks, World Scientific, 2016. doi:10.1142/q0033.
[46] X. Zhang, J. Zhu, Q. Wang, and H. Zhao, "Identifying influential nodes in complex networks with community structure," Knowledge-Based Systems, vol.42, pp.74–84, 2013. doi:10.1016/j.knosys.2013.01.017.
[47] S. Pei, L. Muchnik, J.S. Andrade, Jr, Z. Zheng, and H.A. Makse, "Searching for superspreaders of information in real-world social media," Scientific Reports, vol.4, p.5547, 2014. doi:10.1038/srep05547.
[48] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, and H.A. Makse, "Identification of influential spreaders in complex networks," Nature Physics, vol.6, no.11, pp.888–893, 2010. doi:10.1038/nphys1746.
[49] Q. Li, T. Zhou, L. Lü, and D. Chen, "Identifying influential spreaders by weighted LeaderRank," Physica A: Statistical Mechanics and its Applications, vol.404, pp.47–55, 2014. doi:10.1016/j.physa.2014.02.041.
[50] K. Sugiyama, H. Ohsaki, and M. Imase, "Estimating node characteristics from topological structure of social networks," IEICE Trans. Commun., vol.E92-B, no.10, pp.3094–3101, Oct. 2009. doi:10.1587/transcom.E92.B.3094.
[51] L.C. Freeman, "Centrality in social networks conceptual clarification," Soc. Networks, vol.1, no.3, pp.215–239, 1979. doi:10.1016/0378-8733(78)90021-7.

[52] J. Shetty and J. Adibi, “Discovering important nodes through graph entropy the case of Enron email database,” Proc. 3rd International Workshop on Link Discovery (LinkKDD’05), pp.74–81, 2005. doi:10.1145/1134271.1134282.
[53] S.P. Borgatti, “Identifying sets of key players in a social network,” Computational & Mathematical Organization Theory, vol.12, no.1, pp.21–34, 2006. doi:10.1007/s10588-006-7084-x.
[54] C. Budak, D. Agrawal, and A. El Abbadi, “Limiting the spread of misinformation in social networks,” Proc. 20th International Conf. on World Wide Web (WWW’11), pp.665–674, 2011. doi:10.1145/1963405.1963499.
[55] K. Okamoto, W. Chen, and X.Y. Li, “Ranking of closeness centrality for large-scale social networks,” Lecture Notes in Computer Science, vol.5059, pp.186–195, 2008. doi:10.1007/978-3-540-69311-6_21.
[56] S.P. Borgatti, K.M. Carley, and D. Krackhardt, “On the robustness of centrality measures under conditions of imperfect data,” Soc. Networks, vol.28, no.2, pp.124–136, 2006. doi:10.1016/j.socnet.2005.05.001.
[57] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” Proc. 8th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD’02), pp.61–70, 2002. doi:10.1145/775047.775057.
[58] P. Domingos and M. Richardson, “Mining the network value of customers,” Proc. 7th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD’01), pp.57–66, 2001. doi:10.1145/502512.502525.
[59] D.R. White and S.P. Borgatti, “Betweenness centrality measures for directed graphs,” Soc. Networks, vol.16, no.4, pp.335–346, 1994. doi:10.1016/0378-8733(94)90015-9.
[60] T. Opsahl, F. Agneessens, and J. Skvoretz, “Node centrality in weighted networks: Generalizing degree and shortest paths,” Soc. Networks, vol.32, no.3, pp.245–251, 2010. doi:10.1016/j.socnet.2010.03.006.
[61] L.C. Freeman, S.P. Borgatti, and D.R. White, “Centrality in valued graphs: A measure of betweenness based on network flow,” Soc. Networks, vol.13, no.2, pp.141–154, 1991. doi:10.1016/0378-8733(91)90017-N.
[62] U. Brandes, “On variants of shortest-path betweenness centrality and their generic computation,” Soc. Networks, vol.30, no.2, pp.136–145, 2008. doi:10.1016/j.socnet.2007.11.001.
[63] M.E.J. Newman, “A measure of betweenness centrality based on random walks,” Soc. Networks, vol.27, no.1, pp.39–54, 2005. doi:10.1016/j.socnet.2004.11.009.
[64] S. Dolev, Y. Elovici, and R. Puzis, “Routing betweenness centrality,” J. ACM, vol.57, no.4, pp.25:1–25:27, 2010. doi:10.1145/1734213.1734219.
[65] S.P. Borgatti and M.G. Everett, “A graph-theoretic perspective on centrality,” Soc. Networks, vol.28, no.4, pp.466–484, 2006. doi:10.1016/j.socnet.2005.11.005.
[66] L. Katz, “A new status index derived from sociometric analysis,” Psychometrika, vol.18, no.1, pp.39–43, 1953. doi:10.1007/BF02289026.
[67] P. Bonacich, “Factoring and weighting approaches to status scores and clique identification,” J. Mathematical Sociology, vol.2, no.1, pp.113–120, 1972. doi:10.1080/0022250X.1972.9989806.
[68] P. Hage and F. Harary, “Eccentricity and centrality in networks,” Soc. Networks, vol.17, no.1, pp.57–63, 1995. doi:10.1016/0378-8733(94)00248-9.
[69] S. Brin and L. Page, “The anatomy of a large-scale hypertextual Web search engine,” Computer Networks and ISDN Systems, vol.30, no.1, pp.107–117, 1998. doi:10.1016/S0169-7552(98)00110-X.
[70] S.B. Seidman, “Network structure and minimum degree,” Soc. Networks, vol.5, no.3, pp.269–287, 1983. doi:10.1016/0378-8733(83)90028-X.
[71] S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes, “K-core organization of complex networks,” Phys. Rev. Lett., vol.96, no.4, p.040601, 2006. doi:10.1103/PhysRevLett.96.040601.
[72] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, “A model of Internet topology using k-shell decomposition,” Proc. National Academy of Sciences, vol.104, no.27, pp.11150–11154, 2007. doi:10.1073/pnas.0701175104.
[73] F. Morone and H.A. Makse, “Influence maximization in complex networks through optimal percolation,” Nature, vol.524, no.7563, pp.65–68, 2015. doi:10.1038/nature14604.
[74] X. Teng, S. Pei, F. Morone, and H.A. Makse, “Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks,” Scientific Reports, vol.6, pp.36043:1–36043:11, 2016. doi:10.1038/srep36043.
[75] A. Goyal, F. Bonchi, and L.V. Lakshmanan, “A data-based approach to social influence maximization,” Proc. VLDB Endowment, vol.5, no.1, pp.73–84, 2011. doi:10.14778/2047485.2047492.
[76] Y. Li, W. Chen, Y. Wang, and Z.L. Zhang, “Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships,” Proc. 6th ACM International Conf. on Web Search and Data Mining (WSDM’13), pp.657–666, 2013. doi:10.1145/2433396.2433478.
[77] M. Gomez-Rodriguez, L. Song, N. Du, H. Zha, and B. Schölkopf, “Influence estimation and maximization in continuous-time diffusion networks,” ACM Trans. Inform. Syst. (TOIS), vol.34, no.2, pp.9:1–9:33, 2016. doi:10.1145/2824253.
[78] W. Chen, W. Lu, and N. Zhang, “Time-critical influence maximization in social networks with time-delayed diffusion process,” Proc. Twenty-Sixth AAAI Conf. on Artificial Intelligence (AAAI’12), pp.592–598, 2012.
[79] Y. Tang, Y. Shi, and X. Xiao, “Influence maximization in near-linear time: A martingale approach,” Proc. 2015 ACM SIGMOD International Conf. on Management of Data (SIGMOD’15), pp.1539–1554, 2015. doi:10.1145/2723372.2723734.
[80] W. Chen, C. Wang, and Y. Wang, “Scalable influence maximization for prevalent viral marketing in large scale social networks,” Proc. 16th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD’10), pp.1029–1038, 2010. doi:10.1145/1835804.1835934.
[81] W. Chen, Y. Wang, and S. Yang, “Efficient influence maximization in social networks,” Proc. 15th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD’09), pp.199–208, 2009. doi:10.1145/1557019.1557047.
[82] T. Maehara, H. Suzuki, and M. Ishihata, “Exact computation of influence spread by binary decision diagrams,” Proc. 26th International Conf. on World Wide Web (WWW’17), pp.947–956, 2017. doi:10.1145/3038912.3052567.
[83] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, “Cost-effective outbreak detection in networks,” Proc. 13th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD’07), pp.420–429, 2007. doi:10.1145/1281192.1281239.
[84] A. Goyal, W. Lu, and L.V. Lakshmanan, “CELF++: Optimizing the greedy algorithm for influence maximization in social networks,” Proc. 20th International Conf. on World Wide Web (WWW’11) Companion, pp.47–48, 2011. doi:10.1145/1963192.1963217.
[85] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier, “Maximizing social influence in nearly optimal time,” Proc. Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA’14), pp.946–957, 2014. doi:10.1137/1.9781611973402.70.
[86] Y. Tang, X. Xiao, and Y. Shi, “Influence maximization: Near-optimal time complexity meets practical efficiency,” Proc. 2014 ACM SIGMOD International Conf. on Management of Data (SIGMOD’14), pp.75–86, June 2014. doi:10.1145/2588555.2593670.
[87] H.T. Nguyen, M.T. Thai, and T.N. Dinh, “Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks,” Proc. 2016 International Conf. on Management of Data (SIGMOD’16), pp.695–710, 2016. doi:10.1145/2882903.2915207.

[88] X. Wang, Y. Zhang, W. Zhang, X. Lin, and C. Chen, "Bring order into the samples: A novel scalable method for influence maximization," IEEE Transactions on Knowledge and Data Engineering, vol.29, no.2, pp.243–256, 2017. doi:10.1109/TKDE.2016.2624734.
[89] K. Jung, W. Heo, and W. Chen, "IRIE: Scalable and robust influence maximization in social networks," Proc. 12th IEEE International Conf. on Data Mining (ICDM'12), pp.918–923, 2012. doi:10.1109/ICDM.2012.79.
[90] A. Goyal, W. Lu, and L.V. Lakshmanan, "SIMPATH: An efficient algorithm for influence maximization under the linear threshold model," Proc. 11th IEEE International Conf. on Data Mining (ICDM'11), pp.211–220, 2011. doi:10.1109/ICDM.2011.132.
[91] J. Weng, E.P. Lim, J. Jiang, and Q. He, "TwitterRank: Finding topic-sensitive influential Twitterers," Proc. 3rd ACM International Conf. on Web Search and Data Mining (WSDM'10), pp.261–270, 2010. doi:10.1145/1718487.1718520.
[92] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet allocation," J. Machine Learning Research, vol.3, pp.993–1022, 2003.
[93] E. Bakshy, J.M. Hofman, W.A. Mason, and D.J. Watts, "Everyone's an influencer: Quantifying influence on Twitter," Proc. 4th ACM International Conf. on Web Search and Data Mining (WSDM'11), pp.65–74, 2011. doi:10.1145/1935826.1935845.
[94] S. Mihara, S. Tsugawa, and H. Ohsaki, "Influence maximization problem for unknown social networks," Proc. 2015 IEEE/ACM International Conf. on Advances in Soc. Networks Analysis and Mining (ASONAM'15), pp.1539–1546, 2015. doi:10.1145/2808797.2808885.
[95] S. Mihara, S. Tsugawa, and H. Ohsaki, "On the effectiveness of random jumps in an influence maximization algorithm for unknown graphs," Proc. 31st International Conf. on Information Networking (ICOIN'17), pp.395–400, 2017. doi:10.1109/ICOIN.2017.7899466.
[96] H. Kim, K. Beznosov, and E. Yoneki, "Finding influential neighbors to maximize information diffusion in twitter," Proc. 23rd International Conf. on World Wide Web (WWW'14) Companion, pp.701–706, 2014. doi:10.1145/2567948.2579358.
[97] D. Taylor, S.A. Myers, A. Clauset, M.A. Porter, and P.J. Mucha, "Eigenvector-based centrality measures for temporal networks," Multiscale Modeling & Simulation, vol.15, no.1, pp.537–574, 2017. doi:10.1137/16M1066142.
[98] L.E. Rocha and N. Masuda, "Random walk centrality for temporal networks," New J. Phys., vol.16, no.6, p.063023, 2014. doi:10.1088/1367-2630/16/6/063023.
[99] A. Solé-Ribalta, M. De Domenico, S. Gómez, and A. Arenas, "Centrality rankings in multiplex networks," Proc. 2014 ACM Conf. on Web Science (WebSci'14), pp.149–155, 2014. doi:10.1145/2615569.2615687.
[100] A. Halu, R.J. Mondragón, P. Panzarasa, and G. Bianconi, "Multiplex PageRank," PloS ONE, vol.8, no.10, p.e78293, 2013. doi:10.1371/journal.pone.0078293.
[101] Z. Huang and D.D. Zeng, "A link prediction approach to anomalous email detection," Proc. 2006 IEEE International Conf. on Systems, Man, and Cybernetics (SMC'06), pp.1131–1136, 2006. doi:10.1109/ICSMC.2006.384552.
[102] H.H. Song, T.W. Cho, V. Dave, Y. Zhang, and L. Qiu, "Scalable proximity estimation and link prediction in online social networks," Proc. 9th ACM SIGCOMM Conf. on Internet Measurement (IMC'09), pp.322–335, 2009. doi:10.1145/1644893.1644932.
[103] A. Clauset, C. Moore, and M.E.J. Newman, "Hierarchical structure and the prediction of missing links in networks," Nature, vol.453, no.7191, pp.98–101, 2008. doi:10.1038/nature06830.
[104] W.Q. Wang, Q.M. Zhang, and T. Zhou, "Evaluating network models: A likelihood analysis," EPL (Europhysics Letters), vol.98, no.2, p.28004, 2012. doi:10.1209/0295-5075/98/28004.
[105] L. Lü and T. Zhou, "Link prediction in weighted networks: The role of weak ties," EPL (Europhysics Letters), vol.89, p.18001, 2010. doi:10.1209/0295-5075/89/18001.
[106] T. Murata and S. Moriyasu, "Link prediction of social networks based on weighted proximity measures," IEEE/WIC/ACM International Conf. on Web Intelligence, pp.85–88, 2007. doi:10.1109/WI.2007.71.
[107] S. Tsugawa and K. Kito, "Retweets as a predictor of relationships among users on social media," PloS ONE, vol.12, no.1, p.e0170279, 2017. doi:10.1371/journal.pone.0170279.
[108] D. Schall, "Link prediction in directed social networks," Soc. Netw. Anal. Min., vol.4, no.1, p.157, 2014. doi:10.1007/s13278-014-0157-9.
[109] R.N. Lichtenwalter, J.T. Lussier, and N.V. Chawla, "New perspectives and methods in link prediction," Proc. 16th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD'10), pp.243–252, 2010. doi:10.1145/1835804.1835837.
[110] M.E.J. Newman, "Clustering and preferential attachment in growing networks," Phys. Rev. E, vol.64, no.2, p.025102, 2001. doi:10.1103/PhysRevE.64.025102.
[111] G. Salton and M.J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1986.
[112] L.A. Adamic and E. Adar, "Friends and neighbors on the Web," Soc. Networks, vol.25, no.3, pp.211–230, 2003. doi:10.1016/S0378-8733(03)00009-1.
[113] T. Zhou, L. Lü, and Y.C. Zhang, "Predicting missing links via local information," The European Physical Journal B, vol.71, no.4, pp.623–630, 2009. doi:10.1140/epjb/e2009-00335-8.
[114] Z. Lu, B. Savas, W. Tang, and I.S. Dhillon, "Supervised link prediction using multiple sources," Proc. IEEE 10th International Conf. on Data Mining (ICDM' 10), pp.923–928, 2010. doi:10.1109/ICDM.2010.112.
[115] G. Jeh and J. Widom, "SimRank: A measure of structural-context similarity," Proc. eighth ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD'02), pp.538–543, 2002. doi:10.1145/775047.775126.
[116] H. Kashima, T. Kato, Y. Yamanishi, M. Sugiyama, and K. Tsuda, "Link propagation: A fast semi-supervised learning algorithm for link prediction," Proc. 10th SIAM International Conf. on Data Mining, pp.1099–1110, 2009. doi:10.1137/1.9781611972795.94.
[117] H. Kashima and N. Abe, "A parameterized probabilistic model of network evolution for supervised link prediction," Proc. 6th International Conf. on Data Mining (ICDM'06), pp.340–349, 2006. doi:10.1109/ICDM.2006.8.
[118] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," Proc. Workshop on Link Analysis, Counter-terrorism and Security, 2006.
[119] X. Li, N. Du, H. Li, K. Li, J. Gao, and A. Zhang, "A deep learning approach to link prediction in dynamic networks," Proc. 2014 SIAM International Conf. on Data Mining (SDM'14), pp.289–297, 2014. doi:10.1137/1.9781611973440.33.
[120] M. Pujari and R. Kanawati, "Supervised rank aggregation approach for link prediction in complex networks," Proc. 21st International Conference on World Wide Web Companion, pp.1189–1196, 2012. doi:10.1145/2187980.2188260.
[121] P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: The state-of-the-art," Sci. China Inf. Sci., vol.58, no.1, pp.1–38, 2015. doi:10.1007/s11432-014-5237-y.
[122] T. Tylenda, R. Angelova, and S. Bedathur, "Towards time-aware link prediction in evolving social networks," Proc. 3rd Workshop on Social Network Mining and Analysis (SNAKDD 2009), pp.9:1–9:10, 2009. doi:10.1145/1731011.1731020.
[123] S. Tsugawa and H. Ohsaki, "Effectiveness of link prediction for face-to-face behavioral networks," PloS ONE, vol.8, no.12, p.e81727, 2013. doi:10.1371/journal.pone.0081727.
[124] M. Pujari and R. Kanawati, "Link prediction in multiplex networks," AIMS Networks & Heterogeneous Media Journal, vol.10, no.1, pp.17–35, 2015. doi:10.3934/nhm.2015.10.17.
[125] C. De Bacco, E.A. Power, D.B. Larremore, and C. Moore, "Com-

munity detection, link prediction, and layer interdependence in multilayer networks," Phys. Rev. E, vol.95, no.4, p.042317, 2017. doi:10.1103/PhysRevE.95.042317.
[126] M.E.J. Newman, "Communities, modules and large-scale structure in networks," Nature Physics, vol.8, no.1, pp.25–31, 2012. doi:10.1038/nphys2162.
[127] S. Sahebi and W.W. Cohen, "Community-based recommendations: A solution to the cold start problem," Proc. Workshop on Recommender Systems and the Social Web, p.60, 2011.
[128] J. Leskovec, L.A. Adamic, and B.A. Huberman, "The dynamics of viral marketing," ACM Trans. Web, vol.1, no.1, pp.5:1–5:39, 2007. doi:10.1145/1232722.1232727.
[129] T. Chakraborty, A. Dalmia, A. Mukherjee, and N. Ganguly, "Metrics for community analysis: A survey," ACM Comput. Surv., vol.50, no.4, pp.54:1–54:32, 2017. doi:10.1145/3091106.
[130] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," New J. Phys., vol.11, no.3, p.033015, 2009. doi:10.1088/1367-2630/11/3/033015.
[131] H. Shen, X. Cheng, K. Cai, and M.B. Hu, "Detect overlapping and hierarchical community structure in networks," Physica A: Statistical Mechanics and its Applications, vol.388, no.8, pp.1706–1712, 2009. doi:10.1016/j.physa.2008.12.021.
[132] M. Rosvall and C.T. Bergstrom, "Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems," PloS ONE, vol.6, no.4, p.e18209, 2011. doi:10.1371/journal.pone.0018209.
[133] F.D. Malliaros and M. Vazirgiannis, "Clustering and community detection in directed networks: A survey," Physics Reports, vol.533, no.4, pp.95–142, 2013. doi:10.1016/j.physrep.2013.08.002.
[134] T. Takaguchi and Y. Yoshida, "Cycle and flow trusses in directed networks," Royal Society Open Science, vol.3, no.11, p.160270, 2016. doi:10.1098/rsos.160270.
[135] M. Girvan and M.E.J. Newman, "Community structure in social and biological networks," Proc. National Academy of Sciences, vol.99, no.12, pp.7821–7826, 2002. doi:10.1073/pnas.122653799.
[136] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, "On modularity clustering," IEEE Trans. Knowl. Data Eng., vol.20, no.2, pp.172–188, 2008. doi:10.1109/TKDE.2007.190689.
[137] M.E.J. Newman, "Fast algorithm for detecting community structure in networks," Phys. Rev. E, vol.69, no.6, p.066133, 2004. doi:10.1103/PhysRevE.69.066133.
[138] A. Clauset, M.E.J. Newman, and C. Moore, "Finding community structure in very large networks," Phys. Rev. E, vol.70, no.6, p.066111, 2004. doi:10.1103/PhysRevE.70.066111.
[139] V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Statistical Mechanics: Theory and Experiment, vol.2008, no.10, p.P10008, 2008. doi:10.1088/1742-5468/2008/10/P10008.
[140] H. Shiokawa, Y. Fujiwara, and M. Onizuka, "Fast algorithm for modularity-based graph clustering," Proc. Twenty-Seventh AAAI Conf. on Artificial Intelligence (AAAI'13), pp.1170–1176, 2013.
[141] R. Guimera and L.A.N. Amaral, "Functional cartography of complex metabolic networks," Nature, vol.433, no.7028, pp.895–900, 2005. doi:10.1038/nature03288.
[142] S. White and P. Smyth, "A spectral clustering approach to finding communities in graphs," Proc. 2005 SIAM International Conf. on Data Mining (SDM'05), pp.274–285, 2005. doi:10.1137/1.9781611972757.25.
[143] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," Phys. Rev. E, vol.72, no.2, p.027104, 2005. doi:10.1103/PhysRevE.72.027104.
[144] A.L. Traud, P.J. Mucha, and M.A. Porter, "Social structure of Facebook networks," Physica A: Statistical Mechanics and its Applications, vol.391, no.16, pp.4165–4180, 2012. doi:10.1016/j.physa.2011.12.021.
[145] J.P. Onnela, S. Arbesman, M.C. González, A.L. Barabási, and N.A. Christakis, "Geographic constraints on social network groups," PLoS ONE, vol.6, no.4, p.e16939, 2011. doi:10.1371/journal.pone.0016939.
[146] A. Lancichinetti and S. Fortunato, "Limits of modularity maximization in community detection," Phys. Rev. E, vol.84, no.6, p.066122, 2011. doi:10.1103/PhysRevE.84.066122.
[147] R. Guimera, M. Sales-Pardo, and L.A.N. Amaral, "Modularity from fluctuations in random graphs and complex networks," Phys. Rev. E, vol.70, no.2, p.025101, 2004. doi:10.1103/PhysRevE.70.025101.
[148] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," Proc. National Academy of Sciences, vol.104, no.1, pp.36–41, 2007. doi:10.1073/pnas.0605965104.
[149] B.H. Good, Y.A. de Montjoye, and A. Clauset, "Performance of modularity maximization in practical contexts," Phys. Rev. E, vol.81, no.4, p.046106, 2010. doi:10.1103/PhysRevE.81.046106.
[150] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," Nature, vol.435, pp.814–818, 2005. doi:10.1038/nature03607.
[151] J.M. Kumpula, M. Kivelä, K. Kaski, and J. Saramäki, "Sequential algorithm for fast clique percolation," Phys. Rev. E, vol.78, no.2, p.026109, 2008. doi:10.1103/PhysRevE.78.026109.
[152] Y.Y. Ahn, J.P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," Nature, vol.466, no.7307, pp.761–764, 2010. doi:10.1038/nature09182.
[153] T. Evans and R. Lambiotte, "Line graphs, link partitions, and overlapping communities," Phys. Rev. E, vol.80, no.1, p.016105, 2009. doi:10.1103/PhysRevE.80.016105.
[154] J.J. Whang, D.F. Gleich, and I.S. Dhillon, "Overlapping community detection using neighborhood-inflated seed expansion," IEEE Trans. Knowl. Data Eng., vol.28, no.5, pp.1272–1284, 2016. doi:10.1109/TKDE.2016.2518687.
[155] F. Havemann, M. Heinz, A. Struck, and J. Gläser, "Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels," J. Statistical Mechanics: Theory and Experiment, vol.2011, no.01, p.P01023, 2011. doi:10.1088/1742-5468/2011/01/P01023.
[156] R. Andersen, F. Chung, and K. Lang, "Local graph partitioning using PageRank vectors," Proc. 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06), pp.475–486, 2006. doi:10.1109/FOCS.2006.44.
[157] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.8, pp.888–905, 2000. doi:10.1109/34.868688.
[158] J. Leskovec, K.J. Lang, A. Dasgupta, and M.W. Mahoney, "Statistical properties of community structure in large social and information networks," Proc. 17th International Conf. on World Wide Web (WWW'08), pp.695–704, 2008. doi:10.1145/1367497.1367591.
[159] J. Yang and J. Leskovec, "Overlapping community detection at scale: A nonnegative matrix factorization approach," Proc. 6th ACM International Conf. on Web Search and Data Mining (WSDM'13), pp.587–596, 2013. doi:10.1145/2433396.2433471.
[160] B. Karrer and M.E.J. Newman, "Stochastic blockmodels and community structure in networks," Phys. Rev. E, vol.83, no.1, p.016107, 2011. doi:10.1103/PhysRevE.83.016107.
[161] P. Zhang, C. Moore, and M.E.J. Newman, "Community detection in networks with unequal groups," Phys. Rev. E, vol.93, no.1, p.012303, 2016. doi:10.1103/PhysRevE.93.012303.
[162] E. Abbe, "Community detection and stochastic block models: Recent developments," arXiv preprint arXiv:1703.10146, 2017.
[163] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," Phys. Rev. E, vol.80, no.5, p.056117, 2009. doi:10.1103/PhysRevE.80.056117.
[164] J. Leskovec, K.J. Lang, and M. Mahoney, "Empirical comparison of algorithms for network community detection," Proc. 19th International Conf. on World Wide Web (WWW'10), pp.631–640,

2010. doi:10.1145/1772690.1772755.

[165] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," Knowledge and Information Systems, vol.42, no.1, pp.181–213, 2015. doi:10.1007/s10115-013-0693-z.

[166] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, and J.P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," Science, vol.328, no.5980, pp.876–878, 2010. doi:10.1126/science.1184819.

[167] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," Proc. 2010 International Conf. on Advances in Soc. Networks Analysis and Mining (ASONAM'10), pp.176–183, 2010. doi:10.1109/ASONAM.2010.17.

[168] T. Hartmann, A. Kappes, and D. Wagner, "Clustering evolving networks," in Algorithm Engineering, pp.280–329, Springer, 2016. doi:10.1007/978-3-319-49487-6_9.

[169] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: Data forwarding in disconnected mobile ad hoc networks," IEEE Commun. Mag., vol.44, no.11, 2006. doi:10.1109/MCOM.2006.248176.

[170] M.J. Khabbaz, C.M. Assi, and W.F. Fawaz, "Disruption-tolerant networking: A comprehensive survey on recent developments and persisting challenges," IEEE Commun. Surveys Tuts., vol.14, no.2, pp.607–640, 2012. doi:10.1109/SURV.2011.041911.00093.

[171] H. Hartenstein and L. Laberteaux, "A tutorial survey on vehicular ad hoc networks," IEEE Commun. Mag., vol.46, no.6, pp.164–171, 2008. doi:10.1109/MCOM.2008.4539481.

[172] A.B. MacKenzie and S.B. Wicker, "Game theory in communications: Motivation, explanation, and application to power control," Proc. 2001 IEEE Global Telecommunications Conf. (GLOBECOM'01), pp.821–826, 2001. doi:10.1109/GLOCOM.2001.965533.

[173] M.H. Manshaei, Q. Zhu, T. Alpcan, T. Bacşar, and J.P. Hubaux, "Game theory meets network security and privacy," ACM Comput. Surv., vol.45, no.3, p.25, 2013. doi:10.1145/2480741.2480742.

[174] K. Yamamoto, "A comprehensive survey of potential game approaches to wireless networks," IEICE Trans. Commun., vol.E98-B, no.9, pp.1804–1823, Sept. 2015. doi:10.1587/transcom.E98.B.1804.

[175] B. Gu, K. Yamori, S. Xu, and Y. Tanaka, "Multi-stage non-cooperative game for pricing and connection admission control in wireless local area networks," IEICE Trans. Commun., vol.E96-B, no.7, pp.1986–1996, July 2013. doi:10.1587/transcom.E96.B.1986.

[176] Y. Cao and Z. Sun, "Routing in delay/disruption tolerant networks: A taxonomy, survey and challenges," IEEE Commun. Surveys Tuts., vol.15, no.2, pp.654–677, 2013. doi:10.1109/SURV.2012.042512.00053.

[177] K. Wei, X. Liang, and K. Xu, "A survey of social-aware routing protocols in delay tolerant networks: Applications, taxonomy and design-related issues," IEEE Commun. Surveys Tuts., vol.16, no.1, pp.556–578, 2014. doi:10.1109/SURV.2013.042313.00103.

[178] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "A survey of social-based routing in delay tolerant networks: Positive and negative social effects," IEEE Commun. Surveys Tuts., vol.15, no.1, pp.387–401, 2013. doi:10.1109/SURV.2012.032612.00004.

[179] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: Social-based forwarding in delay-tolerant networks," IEEE Trans. Mobile Comput., vol.10, no.11, pp.1576–1589, 2011. doi:10.1109/TMC.2010.246.

[180] E.M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," Proc. 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'07), pp.32–40, 2007. doi:10.1145/1288107.1288113.

[181] E.M. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant MANETs," IEEE Trans. Mobile Comput., vol.8, no.5, pp.606–621, 2009. doi:10.1109/TMC.2008.161.

[182] A. Mtibaa, M. May, C. Diot, and M. Ammar, "PeopleRank: Social opportunistic forwarding," Proc. 2010 IEEE Conf. on Computer Communication (INFOCOM'10), pp.1–5, 2010. doi:10.1109/INFCOM.2010.5462261.

[183] F. Li and J. Wu, "LocalCom: A community-based epidemic forwarding scheme in disruption-tolerant networks," Proc. 6th Annual IEEE Communications Society Conf. on Sensor, Mesh and Ad Hoc Communications and Networks (SECON'09), pp.1–9, 2009. doi:10.1109/SAHCN.2009.5168942.

[184] A. Socievole, E. Yoneki, F. De Rango, and J. Crowcroft, "Opportunistic message routing using multi-layer social networks," Proc. 2nd ACM Workshop on High Performance Mobile Opportunistic Systems (HP-MOSys'13), pp.39–46, 2013. doi:10.1145/2507908.2507923.

[185] F. Li, L. Zhao, C. Zhang, Z. Gao, and Y. Wang, "Routing with multi-level cross-community social groups in mobile opportunistic networks," Personal and Ubiquitous Computing, vol.18, no.2, pp.385–396, 2014. doi:10.1007/s00779-013-0657-z.

[186] M. Xiao, J. Wu, and L. Huang, "Community-aware opportunistic routing in mobile social networks," IEEE Trans. Comput., vol.63, no.7, pp.1682–1695, 2014. doi:10.1109/TC.2013.55.

[187] K. Zhu, W. Li, and X. Fu, "Smart: A social-and mobile-aware routing strategy for disruption-tolerant networks," IEEE Trans. Veh. Technol., vol.63, no.7, pp.3423–3434, 2014. doi:10.1109/TVT.2014.2298494.

[188] E. Bulut and B.K. Szymanski, "Exploiting friendship relations for efficient routing in mobile social networks," IEEE Trans. Parallel Distrib. Syst., vol.23, no.12, pp.2254–2265, 2012. doi:10.1109/TPDS.2012.83.

[189] A. Socievole, E. Yoneki, F. De Rango, and J. Crowcroft, "ML-SOR: Message routing using multi-layer social networks in opportunistic communications," Comput. Netw., vol.81, pp.201–219, 2015. doi:10.1016/j.comnet.2015.02.016.

[190] Z. Gao, Y. Shi, S. Chen, and Q. Li, "Exploiting social relationship for opportunistic routing in mobile social networks," IEICE Trans. Commun., vol.E98-B, no.10, pp.2040–2048, Oct. 2015. doi:10.1587/transcom.E98.B.2040.

[191] J.W. Chang and C. Chen, "CROP: Community-relevance-based opportunistic routing in delay tolerant networks," IEICE Trans. Commun., vol.E97-B, no.9, pp.1875–1888, Sept. 2014. doi:10.1587/transcom.E97.B.1875.

[192] A. Vahdat, D. Becker, et al., "Epidemic routing for partially connected ad hoc networks," Technical Report CS-200006, Duke University, 2000.

[193] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know thy neighbor: Towards optimal mapping of contacts to social graphs for DTN routing," Proc. 2010 IEEE Conf. on Computer Communication (INFOCOM'10), pp.1–9, 2010. doi:10.1109/INFCOM.2010.5462135.

[194] T. Hossmann, F. Legendre, and T. Spyropoulos, "From contacts to graphs: Pitfalls in using complex network analysis for DTN routing," Proc. 2009 IEEE Conf. on Computer Communication Workshops, pp.1–6, 2009. doi:10.1109/INFCOMW.2009.5072147.

[195] K. Wei, D. Zeng, S. Guo, and K. Xu, "On social delay-tolerant networking: Aggregation, tie detection, and routing," IEEE Trans. Parallel Distrib. Syst., vol.25, no.6, pp.1563–1573, 2014. doi:10.1109/TPDS.2013.264.

[196] P. Hui and J. Crowcroft, "How small labels create big improvements," Proc. 5th Annual IEEE International Conf. on Pervasive Computing and Communications Workshops, pp.65–70, 2007. doi:10.1109/PERCOMW.2007.55.

[197] W. Moreira, P. Mendes, and S. Sargento, "Opportunistic routing based on daily routines," Proc. 2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'12), pp.1–6, June 2012. doi:10.1109/WoWMoM.2012.6263749.

[198] A. Mei, G. Morabito, P. Santi, and J. Stefa, "Social-aware stateless routing in pocket switched networks," IEEE Trans. Parallel Distrib. Syst., vol.26, no.1, pp.252–261, 2015. doi:10.1109/TPDS.2014.2307857.

[199] P. Hui, E. Yoneki, S.Y. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," Proc. 2nd ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture (MobiArch'07), pp.7:1–7:8, 2007. doi:10.1145/1366919.1366929.

[200] N.P. Nguyen, T.N. Dinh, Y. Xuan, and M.T. Thai, "Adaptive algorithms for detecting community structure in dynamic social networks," Proc. 2011 IEEE Conf. on Computer Communication (INFOCOM'11), pp.2282–2290, 2011. doi:10.1109/INFCOM.2011.5935045.

[201] M. Everett and S.P. Borgatti, "Ego network betweenness," Soc. Networks, vol.27, no.1, pp.31–38, 2005. doi:10.1016/j.socnet.2004.11.007.

[202] M. Yamamoto, "A survey of caching networks in content oriented networks," IEICE Trans. Commun., vol.E99-B, no.5, pp.961–973, May 2016. doi:10.1587/transcom.2015AMI0001.

[203] Y. Wu, S. Yao, Y. Yang, T. Zhou, H. Qian, H. Hu, and M. Hamalainen, "Challenges of mobile social device caching," IEEE Access, vol.4, pp.8938–8947, 2016. doi:10.1109/ACCESS.2016.2633485.

[204] A. Vakali and G. Pallis, "Content delivery networks: Status and trends," IEEE Internet Comput., vol.7, no.6, pp.68–74, 2003. doi:10.1109/MIC.2003.1250586.

[205] V. Jacobson, D.K. Smetters, J.D. Thornton, M.F. Plass, N.H. Briggs, and R.L. Braynard, "Networking named content," Proc. 5th International Conf. on Emerging Networking Experiments and Technologies (CoNEXT'09), pp.1–12, 2009. doi:10.1145/1658939.1658941.

[206] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," IEEE Commun. Mag., vol.50, no.7, pp.26–36, 2012. doi:10.1109/MCOM.2012.6231276.

[207] N. Golrezaei, A.F. Molisch, A.G. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," IEEE Commun. Mag., vol.51, no.4, pp.142–149, 2013. doi:10.1109/MCOM.2013.6495773.

[208] X. Wang, S. Leng, and K. Yang, "Social-aware edge caching in fog radio access networks," IEEE Access, vol.5, pp.8492–8501, 2017. doi:10.1109/ACCESS.2017.2693440.

[209] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," IEEE Commun. Mag., vol.52, no.2, pp.131–139, 2014. doi:10.1109/MCOM.2014.6736753.

[210] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," IEEE Commun. Mag., vol.52, no.8, pp.82–89, 2014. doi:10.1109/MCOM.2014.6871674.

[211] W. Gao, G. Cao, A. Iyengar, and M. Srivatsa, "Supporting cooperative caching in disruption tolerant networks," Proc. 31st International Conf. on Distributed Computing Systems (ICDCS'11), pp.151–161, 2011. doi:10.1109/ICDCS.2011.76.

[212] X. Zhuo, Q. Li, G. Cao, Y. Dai, B. Szymanski, and T. La Porta, "Social-based cooperative caching in DTNs: A contact duration aware approach," Proc. IEEE 8th International Conf. on Mobile Adhoc and Sensor Systems (MASS'11), pp.92–101, 2011. doi:10.1109/MASS.2011.22.

[213] T. Le, Y. Lu, and M. Gerla, "Social caching and content retrieval in disruption tolerant networks (DTNs)," Proc. 2015 International Conf. on Computing, Networking and Communications (ICNC'15), pp.905–910, 2015. doi:10.1109/ICCNC.2015.7069467.

[214] M. Taghizadeh and S. Biswas, "Community based cooperative content caching in social wireless networks," Proc. fourteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'13), pp.257–262, 2013. doi:10.1145/2491288.2491318.

[215] K. Zhu, W. Zhi, L. Zhang, X. Chen, and X. Fu, "Social-aware incentivized caching for D2D communications," IEEE Access, vol.4, pp.7585–7593, 2016. doi:10.1109/ACCESS.2016.2618940.

[216] C. Bernardini, T. Silverston, and O. Festor, "Socially-aware caching strategy for content centric networking," Proc. 2014 IFIP Networking Conf., pp.1–9, 2014. doi:10.1109/IFIPNetworking.2014.6857093.

[217] Z. Wang, L. Sun, X. Chen, W. Zhu, J. Liu, M. Chen, and S. Yang, "Propagation-based social-aware replication for social video contents," Proc. 20th ACM International Conf. on Multimedia (MM'12), pp.29–38, 2012. doi:10.1145/2393347.2393359.

[218] H. Hu, Y. Wen, T.S. Chua, J. Huang, W. Zhu, and X. Li, "Joint content replication and request routing for social video distribution over cloud CDN: A community clustering method," IEEE Trans. Circuits Syst. Video Technol., vol.26, no.7, pp.1320–1333, 2016. doi:10.1109/TCSVT.2015.2455712.

[219] Q. He, D. Wu, and P. Khosla, "SORI: A secure and objective reputation-based incentive scheme for adhoc networks," Proc. 2004 IEEE Wireless Communications and Networking Conf. (WCNC'04), pp.825–830, 2004. doi:10.1109/WCNC.2004.1311293.

[220] K. Sato, R. Hashimoto, M. Yoshino, R. Shinkuma, and T. Takahashi, "Incentive mechanism for P2P content sharing over heterogenous access networks," IEICE Trans. Commun., vol.E91-B, no.12, pp.3821–3830, Dec. 2008. doi:10.1093/ietcom/e91-b.12.3821.

[221] B. Gu, C. Zhang, K. Yamori, Z. Zhou, S. Liu, and Y. Tanaka, "Facilitating incentive-compatible access probability selection in wireless random access networks," IEICE Trans. Commun., vol.E98-B, no.11, pp.2280–2290, Nov. 2015. doi:10.1587/transcom.E98.B.2280.

[222] U. Shevade, H.H. Song, L. Qiu, and Y. Zhang, "Incentive-aware routing in DTNs," Proc. IEEE International Conf. on Network Protocols (ICNP'08), pp.238–247, 2008. doi:10.1109/ICNP.2008.4697042.

[223] B.B. Chen and M.C. Chan, "Mobicent: A credit-based incentive system for disruption tolerant network," Proc. 2010 IEEE Conf. on Computer Communication (INFOCOM'10), pp.1–9, 2010. doi:10.1109/INFCOM.2010.5462136.

[224] Z. Ning, L. Liu, F. Xia, B. Jedari, I. Lee, and W. Zhang, "CAIS: A copy adjustable incentive scheme in community-based socially aware networking," IEEE Trans. Veh. Technol., vol.66, no.4, pp.3406–3419, 2017. doi:10.1109/TVT.2016.2593051.

[225] W. Zhi, K. Zhu, Y. Zhang, and L. Zhang, "Hierarchically social-aware incentivized caching for D2D communications," Proc. 2016 IEEE 22nd International Conf. on Parallel and Distributed Systems (ICPADS'16), pp.316–323, 2016. doi:10.1109/ICPADS.2016.0050.

[226] K. Zhu, W. Li, X. Fu, and L. Zhang, "Data routing strategies in opportunistic mobile social networks: Taxonomy and open challenges," Comput. Netw., vol.93, pp.183–198, 2015. doi:10.1016/j.comnet.2015.10.018.

[227] D. Quercia and S. Hailes, "Sybil attacks against mobile users: Friends and foes to the rescue," Proc. 2010 IEEE Conf. on Computer Communication (INFOCOM'10), pp.1–5, 2010. doi:10.1109/INFCOM.2010.5462218.

[228] F. Xing and W. Wang, "Understanding dynamic denial of service attacks in mobile ad hoc networks," Proc. 2006 IEEE Military Communications Conf. (MILCOM'06), pp.1–7, 2006. doi:10.1109/MILCOM.2006.302178.

[229] C. Scholz, M. Atzmueller, A. Barrat, C. Cattuto, and G. Stumme, "New insights and methods for predicting face-to-face contacts," Proc. 7th International AAAI Conf. on Weblogs and Social Media (ICWSM'13), 2013.

[230] C. Scholz, M. Atzmueller, M. Kibanov, and G. Stumme, "Pre-

dictability of evolving contacts and triadic closure in human face-to-face proximity networks," Social Network Analysis and Mining, vol.4, no.1, p.217, 2014. doi:10.1007/s13278-014-0217-1.

[231] C. Scholz, M. Atzmueller, and G. Stumme, "On the predictability of recurring links in networks of face-to-face proximity," Proc. 23rd International Conf. on World Wide Web Companion, pp.637–642, 2014. doi:10.1145/2567948.2579235.

[232] T. Takaguchi, M. Nakamura, N. Sato, K. Yano, and N. Masuda, "Predictability of conversation partners," Phys. Rev. X, vol.1, no.1, p.011008, 2011. doi:10.1103/PhysRevX.1.011008.

[233] A. Miyauchi and Y. Kawase, "What is a network community?: A novel quality function and detection algorithms," Proc. 24th ACM International on Conf. on Information and Knowledge Management (CIKM'15), pp.1471–1480, 2015. doi:10.1145/2806416.2806555.

[234] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," Proc. 3rd International AAAI Conf. on Weblogs and Social Media (ICWSM'09), pp.361–362, 2009.

[235] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," Genome Research, vol.13, no.11, pp.2498–2504, 2003. doi:10.1101/gr.1239303.

[236] J. Leskovec and R. Sosič, "SNAP: A general-purpose network analysis and graph-mining library," ACM Trans. Intell. Syst. Technol. (TIST), vol.8, no.1, p.1, 2016. doi:10.1145/2898361.

[237] N. Eagle and A.S. Pentland, "Reality mining: Sensing complex social systems," Personal and Ubiquitous Computing, vol.10, no.4, pp.255–268, 2006. doi:10.1007/s00779-005-0046-3.

[238] D. Kotz, T. Henderson, I. Abyzov, and J. Yeo, "CRAWDAD dataset dartmouth/campus (v. 2009-09-09)." Downloaded from `https://crawdad.org/dartmouth/campus/20090909`, 2009.

Sho Tsugawa received the M.E. and Ph.D. degrees from Osaka University, Japan, in 2009 and 2012, respectively. He is currently an assistant professor at the Faculty of Engineering, Information and Systems, University of Tsukuba, Japan. His research interests include network science, social network analysis, and computational social science. He is a member of IEEE, ACM, IEICE, and IPSJ.